SUPPORT MATERIAL

Audio Engineering Society Presents

SOUND REPRODUCTION

The Acoustics and Psychoacoustics of Loudspeakers, Rooms and Headphones

Fourth Edition



Floyd E. Toole
with Sean Olive and Todd Welti

A Focal Press Book

ROUTLEDGE

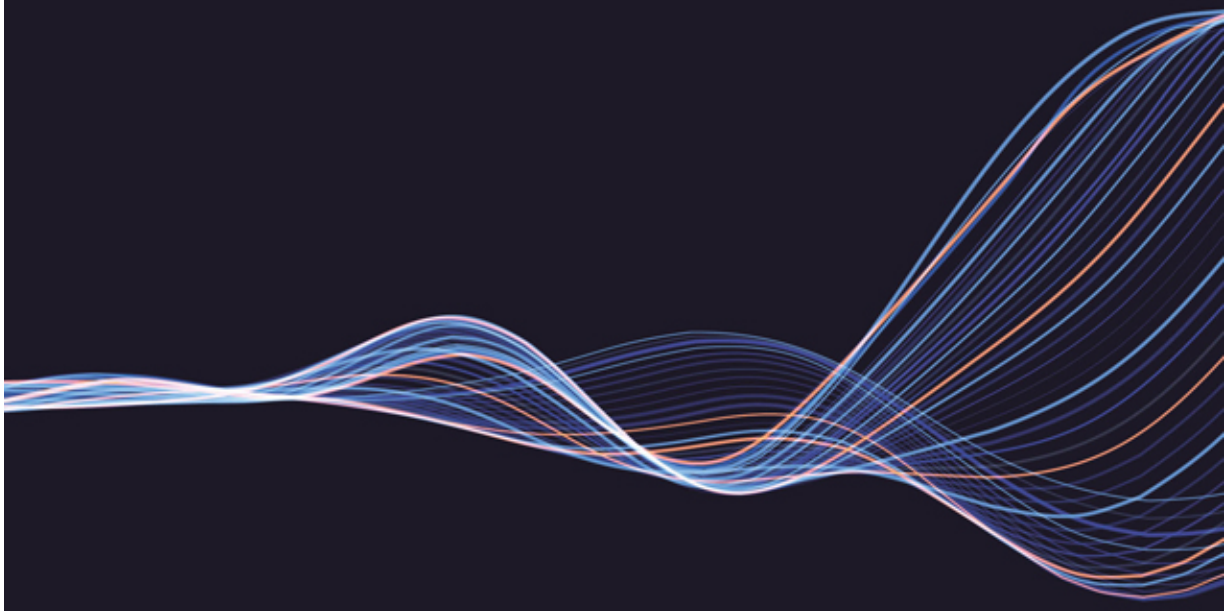
SUPPORT MATERIAL

Audio Engineering Society Presents

SOUND REPRODUCTION

The Acoustics and Psychoacoustics of Loudspeakers, Rooms and Headphones

Fourth Edition



Floyd E. Toole
with Sean Olive and Todd Welti

A **Focal Press** Book

ROUTLEDGE

Sound Reproduction

This fourth edition of *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers, Rooms and Headphones* explains the physical and perceptual processes that are involved in sound reproduction and demonstrates how to use these processes to create high-quality listening experiences in stereo and multichannel formats.

Understanding the principles of sound production is necessary to achieve the goals of sound reproduction in spaces ranging from recording control rooms and home listening rooms to large cinemas. This revised edition includes:

- Sections on 2-channel stereo, headphones, binaural listening and DIY multiple-subwoofer solutions to room resonance problems
- Measurable performance guidelines for loudspeaker designers
- Criteria for the selection of timbrally neutral loudspeakers and headphones using industry-standard measurements
- Guidance for how best to employ loudspeakers in stereo and multichannel formats in small rooms with problematic bass resonances
- Advice on identifying and separating the recording and playback factors that influence sound quality and the spatial/enveloping perceptions that contribute to our satisfaction.

This is essential reading for loudspeaker designers, recording professionals and educators, and aspiring professionals learning in and out of the classroom, as well as being accessible for audio enthusiasts looking to enrich their hobby.

Floyd E. Toole, PhD, is a retired NRC Canada scientist and VP Acoustical Engineering for Harman International. He has won AES Silver and Gold Medals, CEDIA and ALMA lifetime achievement awards, and the IOA Peter Barnett Memorial Award. He is a Fellow of ASA, a past president of AES, and is in the Consumer Technology Hall of Fame.

Sean Olive, PhD, is a senior fellow for Harman International. From 1986 to 1993 he was a research scientist at the National Research Council of Canada. He is a member of the Acoustical Society of America, and a fellow and past president of the Audio Engineering Society.

Todd Welti holds a Master's degree in acoustics from Aalborg University and has over 24 years of research experience at Harman, earning more than 15 patents. A former acoustical consultant, he has contributed to the industry as an author, CEDIA SME, AES committee member and educator. He "retired" in 2025 and currently consults on audio projects that interest him.



Audio Engineering Society Presents …
www.aes.org

Editorial Board

Chair: Francis Rumsey, *Logophon Ltd.*
Hyun Kook Lee, *University of Huddersfield*
Natanya Ford, *University of West England*
Kyle Snyder, *University of Michigan*

Digital Audio Forensics Fundamentals
From Capture to Courtroom
James Zjalic

Drum Sound and Drum Tuning
Bridging Science and Creativity
Rob Toulson

Sound and Recording, 8th Edition
Applications and Theory
Francis Rumsey with Tim McCormick

Performing Electronic Music Live
Kirsten Hermes

Working with the Web Audio API
Joshua Reiss

Modern Recording Techniques, 10th Edition
A Practical Guide to Modern Music Production
David Miles Huber and Emiliano Caballero

Recording Orchestra and Other Classical Music Ensembles, 2nd Edition
Richard King

Sound Reproduction, 4th Edition
The Acoustics and Psychoacoustics of Loudspeakers, Rooms and Headphones
Floyd E. Toole, with Sean Olive and Todd Welti

Classical Recording, 2nd Edition
A Practical Guide in the Decca Tradition
Caroline Haigh, John Dunkerley and Mark Rogers

Handbook for Sound Engineers, 6th Edition
Edited by Glen Ballou and Doug Jones

For more information about this series, please visit: www.routledge.com/Audio-Engineering-Society-Presents/book-series/AES

Sound Reproduction

The Acoustics and Psychoacoustics of Loudspeakers, Rooms and Headphones

Fourth Edition

Floyd E. Toole, with Sean Olive and Todd Welti

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

Designed cover image: SkillUp via Shutterstock

Fourth edition published 2026
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2026 Floyd E. Toole, Sean Olive and Todd Welti

The right of Floyd E. Toole, Sean Olive and Todd Welti to be identified as authors of this work has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks and are used only for identification and explanation without intent to infringe.

First edition published by Focal Press 2008
Third edition published by Routledge 2018

ISBN: 978-1-032-76197-8 (hbk)
ISBN: 978-1-032-76193-0 (pbk)
ISBN: 978-1-003-47749-5 (ebk)

DOI: 10.4324/9781003477495

Typeset in Times New Roman
by Apex CoVantage, LLC

Access the Support Material:
resourcecentre.routledge.com/books/9781032761930

Contents

ACKNOWLEDGMENTS
ABOUT THE AUTHORS

Introduction to the Fourth Edition
FLOYD TOOLE
1 An Overview of the Research
FLOYD TOOLE
1.1 The Scientific Method Applied to Audio
1.2 Fortuitous Events Lead to a Career
1.3 More Chance Factors Steer the Research
1.3.1 The Listening Test That Changed My Career
1.4 Subjective Evaluations Matter, and How They Are Done Matters More
1.5 A Review of Some Measurement Options
1.5.1 Resonances Are Important
1.5.2 Non-linear Distortions
1.5.3 Frequency Response Is Important— But What Is It?
1.5.4 Room Equalization/Calibration Adds Confusion
1.6 Listening in and Through Rooms
1.6.1 Should We Measure the Output of Loudspeakers or the Input to the Ears?
1.7 Circles of Confusion Within the Recording Industry
References
2 A New Standard for Loudspeaker Measurements

FLOYD TOOLE
2.1 Examining the Music Source of the Day: LPs
2.2 The Research Aids Manufacturers and a Broadcaster
2.3 Informative Technical Measurements: Major Progress
2.3.1 Early Results Were Very Revealing
2.4 Learning About the Loudspeaker/Room System
2.5 A View of the Loudspeaker Industry in the 1980s
2.5.1 International Interactions
2.6 Turning Points in Research and Careers
2.7 The “Spinorama” Emerges
2.8 Introducing the Research Team
2.8.1 Sean Olive
2.8.2 Todd Welti
2.8.3 Allan Devantier
2.9 Measurements Have Changed the Loudspeaker Industry
2.9.1 A Real-World Problem for Loudspeaker Users
2.10 A Career Draws to a Close
References
3 Perspectives on Sound Production, Recording and Reproduction
FLOYD TOOLE
3.1 Sound Production
3.2 Sound Reproduction: The Audio Industry
3.2.1 Microphones—The First Transducers
3.2.2 Reproduction Venues
3.3 Can We Reproduce “Reality”? Binaural Techniques
3.3.1 Head Related Transfer Functions: HRTFs
3.3.2 Anecdotal Observations of Adaptation at Work
3.3.3 Adding Credible Visuals Makes a Difference
3.3.4 Binaural Crosstalk-Cancelled Loudspeaker Presentations
3.4 Reproducing the “Perceptual Essence” of Live Performances

3.5 Recording Control and Mastering Rooms—Where “Live” Performances Occur
3.5.1 “Monitor” Loudspeakers Are Not Necessarily “Good” Loudspeakers
3.5.2 An Acoustic Digression in Pro Audio
3.6 The Closed Loop of Cinema Sound
3.7 Summary
References
4 Sound Reproduction: Part One—Mono to Stereo
FLOYD TOOLE
4.1 The Principal Perceptual Variables of Reproduced Sound
4.1.1 Musical Instruments, Microphones and Rooms
4.2 The Monophonic Origins
4.3 Stereo Adds a Channel and Becomes the Standard
4.3.1 The System of Stereo—How It Works
4.3.2 A Phantom Image Problem
4.3.3 Loudspeakers as Stereo Image Stabilizers
4.4 Evaluating Sound Quality and Imaging in Stereo and Mono
4.5 Adding More Channels: A Multichannel Effect
References
5 Loudspeakers, Rooms, Recordings and Adaptation: Key Factors in What and How We Hear
FLOYD TOOLE
5.1 One Room, Two Sound Fields—The Transition Frequency
5.2 Timbral and Spatial Effects Attributable to Rooms
5.3 Resonances Need to Be Heard—Or Not—It Depends
5.3.1 Measuring and Hearing Resonances
5.3.2 Finding and Fixing Resonances: Different Methods—Different Results
5.4 Room Effects, Listening Methods and Adaptation: An Important Experiment

5.5 A Retrospective Examination of These Experimental Results
5.5.1 The Tested Loudspeakers
5.5.2 The Rooms and Loudspeaker/Room Interactions
5.5.3 Program
5.5.4 The Experimental Method
5.6 Adaptation Happens Quickly in Everyday Life
5.7 More Channels, More Pleasure and More Confusion
5.8 Multiple “Rooms” Modify Perceptions
5.8.1 Speculation
5.8.2 Implications for Car Audio
5.8.3 Implications for Headphone Listening
5.8.4 Implications for Low-Frequency Room Resonances
5.9 Conclusion
References
6 The Perception and Measurement of Reflections
FLOYD TOOLE
6.1 A Single Reflection: The Precedence Effect and More
6.1.1 One Reflection—Several Audible Effects
6.1.2 A Real-World View of the Precedence Effect
6.1.3 Reflections From Different Directions
6.2 A Reflection in the Presence of Other Reflections
6.2.1 The “Family” of Curves in a Normal Room
6.3 Effects on Real and Phantom Images
6.4 Results With Music and Other Sounds
6.5 Reflecting, Absorbing and Diffusing Surfaces
6.5.1 Absorbing High Frequencies Is Not Enough
6.5.2 Reflections From Smooth vs. Irregular Surfaces
6.6 Interpreting the Measured Evidence of Reflections
6.6.1 Comb Filtering
6.6.2 Energy-Time Curves

References
7 Space: A Motivator for More Channels
FLOYD TOOLE
7.1 Space Is Fundamental: Reflections Are Important
7.1.1 Perceptually Preferred Amounts of Reflected Sound
7.1.2 Perceptually Preferred Directions for Reflections
7.1.3 Perceptions of Space and Sound Quality Compared
7.2 How Many Channels Do We Need for Envelopment?
7.3 Elevation: The Missing Dimension
References
8 Audio Formats: Part Two—Multichannel Audio
FLOYD TOOLE
8.1 Quadraphonics: Stereo Times Two—A Flawed Idea
8.1.1 Dolby Adapts the Technology to a Better Purpose
8.2 Discrete Multichannel Arrives
8.2.1 The Basic Functions Required of Multichannel Systems
8.2.2 The Subwoofer Channel: LFE and Bass Management
8.3 How Many Channels, Where and What For?
8.3.1 Immersive Audio Arrives
8.4 Multichannel Audio System Options
8.4.1 Channel Numbering Schemes
8.4.2 The Basics of System Layout
8.4.3 Dolby Atmos Music
8.5 Selecting the Appropriate Loudspeakers
8.5.1 Loudspeaker Directivity Requirements
8.5.2 Selecting Front, Left, Center and Right Loudspeakers
8.5.3 Selecting Surround Loudspeakers
8.5.4 A “Physics” Problem With Side Surrounds
8.5.5 Acoustical Treatments Can Help or Hinder
8.5.6 Domestic Room Systems vs. Custom Home Theaters

8.6 The Present and Possible Future of Multichannel Audio
8.6.1 Upmixing, Downmixing and Rendering
8.6.2 A Radical Upmixing Alternative: Create Your Own Concert Hall
8.6.3 Immersive Music Mixing Is Still Evolving
8.6.4 There Is More to Come: Next Generation Audio
References
9 The Perceptual and Physical Dimensions of Sound
FLOYD TOOLE
9.1 The Frequency Domain
9.2 The Amplitude Domain
9.2.1 Absolute and Relative Sound Levels
9.2.2 Different Interpretations of Sound Pressure Levels
9.3 Amplitude and Frequency Together: Frequency Response
9.4 Amplitude and Frequency Together: Equal-Loudness Contours
9.4.1 Loudness Controls: Do They Work?
9.4.2 Loudness, Tone and Tilt Controls— A Summary
9.5 The Boundaries of What We Can Hear
9.5.1 What Is Acceptable Background Noise?
9.6 Linear Distortions: Amplitude, Phase and Polarity
9.6.1 Spectral Deviations From Flat and Smooth: Resonances
9.6.2 Resonances Viewed in Amplitude, Frequency and Time: Waterfalls
9.6.3 Critical Bands, ERBNs and the "Resolution" of the Hearing System
9.6.4 The Audibility of Phase Shift and Group Delay
9.6.5 Phase Shift at Low Frequencies—A Special Case
9.6.6 The Audibility of Absolute Polarity—Which Way Is "Up"?
9.7 Non-linear Distortion
9.8 Dynamic Range and Power Compression
9.9 Wavelength: The Key to Understanding Much in Audio

9.9.1 The Effects on Loudspeaker Directivity and Sound Quality
9.9.2 Acoustical Resonances in Rooms and Enclosures
9.9.3 Understanding Resistive and Membrane Absorbers
9.9.4 Diffusers and Other Sound-Scattering Devices
References
10 Subjective Measurements: Turning Opinion Into Fact
FLOYD TOOLE
10.1 Seeing Is Believing: An Audio Placebo Effect?
10.1.1 Why Blind Listening Is Necessary
10.1.2 Hearing Ability and Listening Performance
10.2 Loudspeaker Reviewing Is a Challenge Few Can Meet
10.3 Designing Listening Tests
10.3.1 Is Stress a Factor? Relax
10.3.2 Stereo vs. Monophonic Listening
10.3.3 The Listening Room
10.3.4 Solutions for the Loudspeaker Position Problem
10.3.5 Solutions for Multiple Listener Tests
10.3.6 Making the Test Blind and Double-Blind
10.3.7 Binaural Record/Replay Loudspeaker Comparisons
10.3.8 Equal-Loudness Testing at a Known Loudness Level
10.3.9 Choosing Program Material
10.3.10 Electrical and Electronic Decisions
10.3.11 Listener Choice and Training
10.3.12 Cultural and National Biases
10.4 Loudness and Sound Level Measurements
10.5 Is It Preference or Accuracy That Is Evaluated?
References
11 Closing the Loop: Predicting Listener Preferences From Measurements
FLOYD TOOLE
11.1 The Olive Experiments—Part One

11.2 The Olive Experiments—Part Two
11.3 The Olive Experiments—Part Three
11.4 Summary
References
12 Room Acoustics and Acoustical Devices
FLOYD TOOLE
12.1 Reverberation Time
12.1.1 Measuring Reverberation Time
12.1.2 Calculating Reverberation Time
12.1.3 Is There a More Useful Metric for Our Purposes?
12.1.4 Diffusion in Small Rooms
12.2 Sound Propagation Basics
12.2.1 Point Sources
12.2.2 Near and Far Fields Influence How We Do Measurements
12.2.3 Line Sources
12.2.4 How Various Sources Interact With a Floor
12.2.5 Are Floor Reflections Special?
12.3 Propagation in Rooms With and Without Reflections
12.3.1 Classic Concert Hall Acoustics
12.3.2 Sound Propagation in Typical Listening Rooms
12.3.3 The Meaning of Steady-State Room Curves
12.3.4 One Loudspeaker in Many Rooms
12.3.5 The “Harman” Room Curve Is a Result, Not a Target
12.3.6 Personal Preferences for Spectral Balance
12.3.7 High Frequencies: Air Absorption and Screen Loss
12.4 Absorbers, Bass “Traps” and Diffusers
12.4.1 Porous/Resistance Absorbers
12.4.2 Diaphragmatic/Membrane Absorbers: Bass “Traps”
12.4.3 Geometric-Shape Diffusers
12.4.4 Engineered-Surface Diffusers

12.5 Loudspeakers and Rooms Determine “Translation”
12.5.1 A Perspective on the “Translation” of Recordings
12.5.2 Close Listening Is Useful, But …
12.5.3 Translation Viewed From the Reproduction End
12.5.4 Translation Recommendations
12.5.5 An Interesting Experiment: Blackbird Studio C
12.6 The Acoustical Treatment of Listening Rooms
12.6.1 Estimating Reverberation Time
References
13 Acoustics: Adjacent-Boundary Effects
FLOYD TOOLE
13.1 The Effects of Solid Angles on Sound Radiated by Point Sources
13.2 Classic Adjacent-Boundary Effects
13.3 Alleviating Adjacent-Boundary Effects
13.3.1 Positioning the Woofer
13.3.2 Equalization
13.4 Loudspeaker Mounting Options and Effects
13.4.1 An Example of Adjacent-Boundary Interference
13.5 Boundary-Friendly Loudspeaker Designs
13.6 Listeners Also Have Boundaries
References
14 Below the Transition Frequency: Optimizing the Experience
TODD WELTI AND FLOYD TOOLE
14.1 Objectives and Assumptions
14.2 The Basics of Room Resonances and Standing Waves
14.3 A More Complete Understanding
14.3.1 Modal Coupling Factors
14.3.2 Modal Resonances
14.3.3 Simplifying a Complex Situation
14.3.4 Tools for Modal Manipulation—A Preview

14.4 Generalized Optimization of Room Dimensions— Does an "Ideal" Room Exist?
14.4.1 Necessary Assumptions
14.4.2 Are Non-rectangular Rooms the Answer?
14.5 Solutions for the Real World
14.5.1 Deliver Energy to the Modes and Dissipate Some of That Energy With Absorbers
14.5.2 Deliver Energy to the Modes and Reduce the Coupling of That Energy to the Listener by Optimizing the Listening Location
14.5.3 Reduce the Energy Delivered to a Bothersome Mode by Optimizing the Loudspeaker/Subwoofer Location
14.5.4 Reduce the Energy Delivered to Bothersome Modes by Using Combined Coupling of Two Sources to Mode
14.5.5 Use Standardized Subwoofer Configurations for Generalized Multiple Subwoofer Mode Manipulation—Basic
14.5.6 Use Standardized Subwoofer Configurations for Generalized Multiple Subwoofer Mode Manipulation—Advanced
14.5.7 Manipulate Modes Using Multiple Subwoofers and Computer Optimized Signal Processing—Rectangular Rooms
14.5.8 Manipulate Modes Using Multiple Subwoofers and Computer Optimized Signal Processing—All Rooms
14.5.9 Final Equalization at Low Frequencies: Global Correction of System Response Without Modal Modification
14.5.10 Revisiting Room Resonances in Time and Space
14.5.11 Do We Hear the Spectral Bump, the Temporal Ringing or Both?
14.5.12 Stereo Bass: Little Ado About Even Less
14.5.13 Bass Management Makes It All Possible
14.6 Summary and Discussion
References

15 Headphones: The Perception and Measurement of Their Sound Quality

SEAN OLIVE

15.1 The Psychoacoustics of Headphones

15.1.1 Spatial Impressions

15.1.2 Timbral Impressions

15.1.3 Reduced Sensitivity to Resonances

15.1.4 Lack of Tactile Experience in Headphones

15.1.5 Immersive Audio Over Loudspeakers vs. Binaural Headphones

15.1.6 Single Stimulus Comparisons, Auditory Memory and Adaptation

15.1.7 Summary

15.2 Headphones: The Primary Devices Through Which We Listen to Reproduced Sound

15.3 Types of Headphones

15.3.1 Headphones With Headbands

15.3.2 Headphones Without Headbands

15.3.3 Headphone Type and Hearing Loss

15.4 Headphone Test Fixtures

15.4.1 Type 3.3 Ear Simulator

15.4.2 Type 4.3 Ear Simulator

15.4.3 Head and Torso Simulators

15.4.4 Headphone Measurement Challenges

15.4.5 Effect of Test Fixture on Headphone Measurements

15.4.6 Comparing Headphones Measured on Humans vs. Test Fixtures

15.4.7 Reproducibility of Headphone Measurements on Humans

15.4.8 The Hyperion Project

15.4.9 Measuring ANC Performance and Transparency in Headphones

15.5 A History of Preferred Headphone Calibrations: Free, Diffuse and Semi-reflective Fields
15.5.1 Free-Field Calibration
15.5.2 Diffuse-Field Calibration
15.5.3 The Semi-reflective Field (SRF) Calibration
15.6 The Harman Research (2012–2024)
15.6.1 Headphone Listening Test Methods: Physical Substitution
15.6.2 The Virtual Headphone Method
15.6.3 The Harman Headphone Target Curve
15.6.4 Testing Harman Target Curve Against Three Highly Regarded Headphones
15.6.5 Listeners' Preferred Levels of Bass and Treble
15.6.6 Benchmarking the Harman Target Against 31 Models
15.6.7 Segmentation of Listeners Based on Preferred Headphone Sound Profile
15.6.8 Predicting Listener Headphone Preferences Based on Objective Measurements
15.6.9 Harman IE Headphone Target for B&K 5128
15.6.10 Proper Application of the Harman Research
15.7 No Correlation Between Price and Frequency Response or Predicted Sound Quality
15.8 Harman Target Tested Against Other Proposed Targets
15.8.1 Sonarworks
15.8.2 Knowles IE Headphone Target Curve
15.8.3 SoundGuys
15.8.4 Rtings
15.8.5 SenseLab Study
15.8.6 Headphones.com: Preference Bounds and DF Calibration
15.9 The Three Faces of Harman Targets
15.10 Headphones for Spatial Audio
15.10.1 Growth in Immersive Audio

15.10.2 Is There a Different Headphone Target for Spatial Audio?
15.11 Topics for Future Research
15.12 Conclusions
References
16 Hearing Loss and Hearing Conservation
FLOYD TOOLE
16.1 Occupational Noise Exposure Limits
16.1.1 What Can Audio Professionals Do for Protection?
16.2 Non-occupational Noise Exposure
16.3 Hidden Hearing Losses
16.4 What Is Enjoyment? It Might Depend on Speech Clarity
16.5 Obsession Can Be a Good Thing
References
17 Summary and Conclusions
FLOYD TOOLE
17.1 Interpreting Anechoic Data
17.2 Looking Back 60 Years
17.3 Where Are We Now?

INDEX

Acknowledgments

Floyd E. Toole: I am grateful to my parents for setting a high moral standard, and for seeing that I got an education. I especially thank my father for teaching me many of his consummate handyman skills, and for being a trusted friend. My wife Noreen Toole shared 58 years of my life. They were very loving, good years, but sadly she succumbed to illness in 2019, 2 days after my father passed away at 107 years. By the greatest of good fortune, I reconnected with a longstanding family friend, Anne Brammer, and at ages 81 and 82, in 2020 these young-at-heart oldsters started a new life together. We have relocated from California to Ottawa, Canada, our home for many years, and are again among family, friends and four seasons. Anne has been very patient with the long process of writing this book, while attending to my health needs. I am a lucky man.

Much of the research described in this book exists only because of the funding and comprehensive technical facilities provided by the National Research Council of Canada (26 years) and Harman International Industries, Inc. (28 years). My colleagues and I investigated the physical and subjective factors involved in sound reproduction and had the privilege of publishing the research findings for the world to see. Sidney Harman and my corporate management colleagues deserve credit for believing in scientific research and investing in it. Harman engineering staff generously provided many of the measurements seen here. The numerous researchers around the world whose efforts have contributed to the science of audio and to the contents of this book have both my respect and my gratitude. Of course, I am indebted to my friends and co-authors who, in addition to their

chapters, have contributed to the scientific content of the book and offered constructive commentary during its preparation.

Sean Olive: My 40-year career in audio research began meeting Floyd Toole on a McGill University field trip to the National Research Council in Ottawa back in 1985. He introduced me to the field, gave me my first and second positions in psychoacoustic research, and has been a constant supportive and generous mentor, colleague and good friend throughout. I am forever indebted to Floyd for introducing me to this field and giving me the opportunity to work and contribute to its knowledge including my small contribution to this book.

I am grateful to Harman for supporting our research over these past 32 years and allowing us to publish the research so that the entire audio industry can benefit from it. Todd Welti has been a constant colleague and friend throughout most of that time which has led to many interesting discussions and discoveries. Along with Todd, I thank the many talented engineers at Harman who contributed to this research as well.

Finally, I thank my children Conor and Mariah and my partner Firuze for their patience and support in my obsession and pursuit of what makes audio sound good.

Todd Welti: I credit my parents for instilling in me a sense of curiosity about the world and a hunger for truth. They encouraged me to get a good education and not rest on my laurels. I thank my first mentor in acoustics, Michael Yantis, and my next two mentors, also my current co-authors and friends Floyd Toole and Sean Olive. Harman International has encouraged our research efforts these many years, and that should be acknowledged. I thank Allan Devantier for assisting in the work on modal optimization in the early 2000s, and Etienne Rivet and Adrian Celestinos for many interesting conversations. Lastly, I thank my wife for giving me space to work some long hours on this.

About the Authors

Floyd E. Toole was a research scientist at the National Research Council of Canada for 26 years, the corporate vice president of acoustical engineering for Harman International for 16 years and a consultant to Harman for 12 years. He has received the Audio Engineering Society Silver and Gold Medal awards, lifetime achievement awards from CEDIA and ALMA International, and the Peter Barnett Award from the Institute of Acoustics (UK). He is a fellow of the Acoustical Society of America, CEDIA and a fellow and past president of the AES. He is in the Consumer Technology Hall of Fame.

Sean E. Olive is an independent consultant and former Senior Fellow in Acoustic Research for Harman International where he has worked since 1993 to 2025. From 1986 to 1993 he was a research scientist at the National Research Council of Canada. He is a member of the Acoustical Society of America and a fellow and past president of the Audio Engineering Society. He received the ALMA International Titanium Driver Award for scientific contributions to the loudspeaker and headphone industry, two AES publication awards and the Harman External Leadership Award.

Todd S. Welti is Distinguished Engineer, Corporate Research group at Harman International. Originally an acoustical consultant for nine years, he worked on projects including environmental, noise control, room acoustics and sound systems. After earning his master's degree in acoustics in 2000 (Aalborg University), he started at Harman, working on many diverse projects including speaker/room interactions, binaural sound, subjective analysis and headphone target curve research. He was worked extensively

with the AES and CEDIA in many capacities over the years and holds over 20 patents granted or pending.

Introduction to the Fourth Edition

Floyd Toole

DOI: 10.4324/9781003477495-1

Reproduced sound is an inescapable part of our lives. Much of it is in the background; occasionally we get selective and bring it into the foreground. For more serious entertainment we sit down and make an event of it, sometimes with video accompaniment. We listen to loudspeakers in homes and cars, headphones, and portable audio devices of many shapes and sizes. Viewed objectively, the sound we hear embraces a wide range of sound qualities. The basic elements of the programs survive many kinds of auditory insults, but, when one hears something better, it is not difficult to appreciate that there is more to the message than was being heard. Most people lack the interest or budget to upgrade their audio delivery devices, and they adapt to compromised circumstances. Life goes on.

But suppose that one appreciates good sound, and wants the best in sound quality, soundstage and imaging in stereo music, and the engaging envelopment and directional diversity that multichannel audio can bring. Suppose one wants to experience the sense of immersion, of being surrounded by an audience at a concert, or the exciting plot-driven special

effects in movies. With enough money we know these are achievable goals, but for many that has been a deterrent.

Fortunately, because of today's technology, skilled engineers and scientific guidance the point of diminishing returns has fallen dramatically in price. Knowledge has made good sound quality available at affordable prices. The entry-level loudspeakers may not be stylish, have deep bass or be able to play at cinema sound levels, but when well designed, they deliver fundamentally uncolored, neutral sound. Add a subwoofer and there can be pleasant surprises. Well-chosen inexpensive headphones can compete with many expensive models in both quality and quantity of sound.

This is one of the most satisfying rewards of the applied science discussed in this book. When research yields knowledge of the relationship between what we hear and what we measure, it is possible to set measurable design objectives for engineers to design products that have a high probability of pleasing listeners. Using the term "pleasing" suggests that the listeners being satisfied are not necessarily as critical as experienced or professional listeners. Interestingly, about 60 years of double-blind, equal-loudness listening tests using listeners of all backgrounds have shown that most listeners with relatively normal hearing agree on what sounds "good", and what sounds "good" has turned out to be the least colored, most neutral sound. Loudspeakers that are s*ubjectively* highly rated turn out to be the most *technically* accurate, exhibiting the least evidence of resonances, the most uniform frequency responses with the widest bandwidth. The correlation between what is heard and what is measured is impressively high, and the relationships make sense.

This is all encouraging, but having "perfect" loudspeakers does not guarantee "perfect" sound. Listening rooms have resonances that cause bass booms and coloration. The bass is different in different seats. Bass accounts for about 30% of the factors contributing to sound quality judgments. It cannot be ignored. The problems are different for every room, and the only solution is to address it in the room after the system is set up. My co-author Todd Welti will explain how.

Headphone listening avoids the room problem entirely but adds other challenges. My other co-author Sean Olive has done much of the existing research on understanding the psychoacoustic relationships between what is measured and heard. Different headphone designs, different measurement fixtures and variations in human anatomy all are complications. Added to these are the fact that stereo recordings are created for loudspeaker reproduction, and what listeners hear in headphones is very different. They are perhaps the most used delivery device for sound reproduction, so it is fortunate that it is now possible to identify good sounding headphones using measurements, as will be discussed in a separate chapter.

A universal issue is that recordings, videos and movies are not all uniformly good. There is a longstanding concern in the industry that the recorded arts do not always “translate” to customers as they should. If consumers are to hear what was created in recording control and mastering rooms, there must be some basic agreement about sound reproduction technology on both the consumer and professional sides of the audio industry. I have called this the “circle of confusion”. It is a real problem, but, as will be found in this book, there are solutions. We now know how to measure and design loudspeakers and headphones that are essentially neutral, transparent sound reproducers. The information is in the public domain and manufacturers are responding.

If what we hear from a particular source is not to our liking, that is what tone controls and accessible equalization are for. But starting with neutral loudspeakers and headphones allows one to experience different spectral balances and to return to neutrality at will.

One can equip a home theater or recording control room to deliver cinema sound levels, or one can set up a stereo or multichannel system that delivers timbral accuracy but that respects neighbors in an apartment or attached dwelling. Making the choice possible is the increasing availability of informative measurements that, when sensibly interpreted, describe how loudspeakers are likely to sound. Elaborate industrial design, exotic woods, power output, directivity, price and size are choices, but the essential sound quality, the timbral accuracy, can be delivered in many forms. This

knowledge benefits the consumer at the point of enjoying the recorded arts, and it benefits those who create the art, the recording engineers who capture and craft the sounds that musicians produce. They all reside within the same audio industry.

There is no doubt that when the best of sound recording and reproduction combine, we can experience long-term satisfaction along with occasional impressive physical and emotional experiences. The solutions exist. Sound quality in loudspeakers and headphones has improved.

> The objective of this book is to maximize the probability of memorable listening experiences in stereo and multichannel music and movies.

Achieving this goal requires a recognition of some fundamental truths about the world of audio. It is a complicated mixture. Figure 0.1 is an attempt to summarize the basic elements and how they relate to one another.

At the top of the illustration, ART is shown to be loosely coupled to SCIENCE. These are aspects of C.P. Snow's "Two Cultures" that often have difficulty communicating and understanding the other's point of view (https://en.wikipedia.org/wiki/The_Two_Cultures).

However, as it has turned out, scientifically conducted double-blind subjective evaluations yield OPINIONS that have some of the properties of FACTS in that they are repeatable across time and different populations of listeners. Therefore, the illustration shows them "somewhat" connected. It is the recognition of these relationships that has allowed the research reported in this book to be successful.

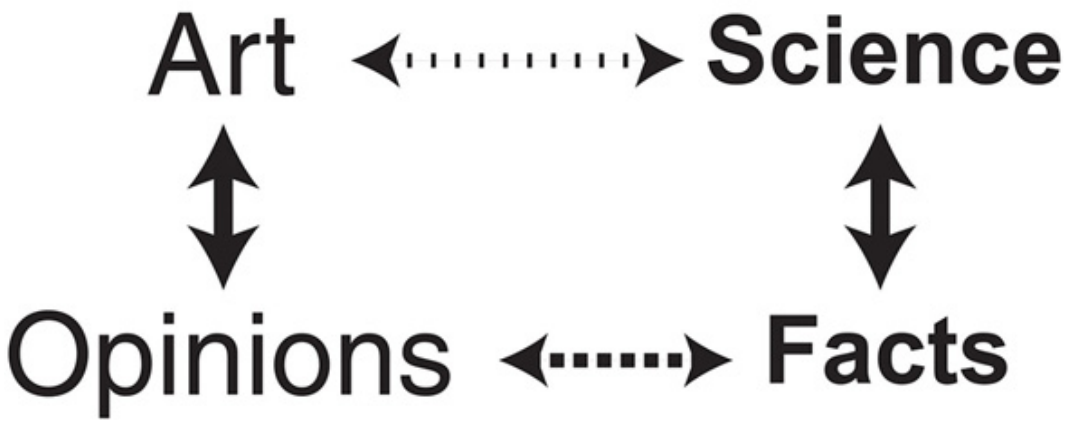


FIGURE 0.1 *This illustration acknowledges the two distinct domains of audio. On the left is the ART (music and movies), which is closely coupled to OPINIONS that people have of different aspects of the art. On the right is SCIENCE, research that attempts to discover relationships between listener opinions about sound, and measurements on the devices that capture, store and deliver the sound. Scientific research, if successful, is closely coupled to FACTS.* ⏎

All evaluations yield opinions, but not all opinions are answers to the question being asked.

This book describes research that investigated patterns in listener opinions about sound quality, observing what influences them and how they relate to one another and to measurable facts. Whatever can be learned about the meaning of measurements must come from guidance provided by subjective evaluations. In this book there are many graphs of measured data that contain information about potential sound quality, but the reality is that:

Listening is the final arbiter of quality—but there are complications.

The major issue is that in the most common listening situations humans are not reliable judges of sound quality. We easily form opinions, but they are influenced by factors that have nothing to do with the sound itself—a form of the "placebo effect". How a product looks, its price, size, reviewers' comments, internet chat and more all lead to expectations that we simply cannot ignore. Musicians, professional audio engineers and product reviewers, despite their experience, are also susceptible to these biases. Peer pressure, marketing influences and often the ego are involved as their product choices reflect their professional judgment.

Many people are unsure of their abilities to judge sound, some thinking they may be "tone-deaf", or just not caring. These insecurities are understandable. However, the reality is very different. Blind and double-blind, equal-loudness, multiple-loudspeaker subjective evaluations have revealed that hidden behind countless listener opinions is a widespread ability to recognize excellence. Some methods of evaluation are much more revealing than others, explaining why there are so many examples of contradictory opinions. Having multiple loudspeakers to instantly compare with one another is a remarkably powerful experimental method, allowing listeners to separate the changing timbral contributions of the loudspeakers from those of the room and those inherent in recordings. Unfortunately, the widely used "take-it-home-and-listen-to-it" technique is the least trustworthy and revealing. It is widely used by reviewers, who, especially if they lack access to meaningful measurements, are susceptible to many biasing factors.

Whether the subjective ratings are described in terms of "preference" or "fidelity", the result is the same; the sounds judged to be closest to uncolored, neutral reproduction receive the highest ratings. The listeners have no knowledge of how a recording *should* sound, but it turns out that most listeners instinctively recognize how a recording *should not* sound. Presumably they are responding to characteristics of reproduced sound that are not "normal" for live sounds. In their comments they describe audible flaws in great detail and sometimes colorful language, but excellence is usually rewarded with a few simple compliments. The "best" loudspeaker,

by this criterion, is the “least bad” one. It seems that “neutrality” is the goal —nothing to complain about.

Listeners of all backgrounds are able to do this, but those with training, experience and normal hearing do it more rapidly, with better consistency. In the end, the product rankings for most listeners are very similar. Obviously, people with hearing loss have a handicap. Their opinions are more variable, and they exhibit biases depending on the nature of their disability. They still enjoy music and movies, but their opinions should be their own. Not surprisingly, age is a factor in this, as well as occupational or recreational exposure to high sound levels. Abuses in widespread headphone listening appears to have resulted in increased hearing problems in young people.

Loudspeakers reproduce whatever is delivered to them by the program source and the upstream electronics. Fortunately, competently designed electronics are essentially transparent, in spite of sometimes creative marketing claims and reviews. That said, there are some technical requirements. Power amplifiers need to be able to drive loudspeaker impedances that routinely drop below the specified value, and not clip on crescendos. Most tube amplifiers change the sound of loudspeakers in measurable and audible ways. As loudspeakers become more neutral, this is an unfortunate situation. LPs are a nostalgic throwback, with artistic covers and user involvement as attractions, but in terms of sound quality they are not capable of presenting “master tape” sound quality, as uncompressed digital sources can.

Sound quality, timbral accuracy, is an essential starting point, but stereo soundstage and imaging, and multichannel and immersive spatial illusions are equally important. This book examines how we perceive sound quality, direction and space in these formats. Occasionally, what we complain about is a fundamental flaw in the recording format, not the playback apparatus. It is important to understand the underlying mechanisms, so we know where audible faults originate and who or what to blame.

The research described in this book and its earlier editions has influenced the loudspeaker industry. The spinorama form of anechoic data presentation

is now embodied in an industry standard. The growing availability of this and other reliably descriptive anechoic measurements on the internet has enabled knowledgeable customers to identify potentially neutral-sounding loudspeakers without ever hearing them. Customers no longer need to rely on traditional manufacturer's specifications, which have a long history of being unreliable.

With few opportunities to listen under reasonable circumstances, this is good news for consumers. Loudspeakers are getting better, and equalization *based on the anechoic data* can elevate many "good" loudspeakers into the "excellent" category. Some internet reviewers with anechoic measurement capability are providing accurate data, and advice for improving the neutrality of loudspeakers. This procedure should not be confused with the "room equalization/calibration" process and the features that exist in numerous consumer products. These are often misused but can offer limited benefits if properly managed. They will be discussed.

This edition has been substantially reorganized and rewritten. It is more biographical than earlier editions. The facts are essential, but understanding how they were discovered, tested and confirmed addresses some concerns expressed by audio enthusiasts and professionals. The treatment is non-mathematical except for explaining the beneficial manipulations of small room resonances. The design of state-of-the-art transducers and loudspeaker systems are computationally intense mechanical and electrical engineering exercises. However, defining the performance targets for loudspeaker designs is an exercise in graphical interpretation. The final judge of success is a human listener, who is listening to recordings made by other humans listening to different loudspeakers under very different circumstances. The "translation" of the recorded art from creator to consumer is a fundamentally important topic that is addressed in some detail.

Several people have contributed significantly to the research that I have initiated or have been involved in at the National Research Council of Canada and at Harman International, including Peter Schuck and Allan Devantier. Their contributions and others are noted and referenced. Sean

Olive and Todd Welti, two of my long-term research collaborators, are co-authors of this book.

We hope you enjoy our efforts.

The content of this book is fixed at publication, but there is supplemental information in the accompanying online content. That content will change with time. In the text, the online content available on the Routledge Resource Centre will be identified by (OR). It can be accessed at: www.routledge.com/9781032761930

CHAPTE
R 1

An Overview of the Research

Floyd Toole

DOI: 10.4324/9781003477495-2

1.1 THE SCIENTIFIC METHOD APPLIED TO AUDIO

Opinions are plentiful in the field of audio. Personal likes and dislikes are assumed. We are all different in many ways, hearing included. For decades it has been assumed that choosing a loudspeaker must be a subjective choice, finding a match to a person's perceptual characteristics and preferences in sound quality. What has been largely underestimated in this scenario is that such subjective evaluations are almost always "sighted". Listeners see the physical loudspeakers, the brand name, they may know the price, perhaps they have read reviews or heard opinions from respected people. All of these factors influence the listener in some way or other, so that the final opinion is not only about the sound they hear. What is also largely underestimated is that humans are very good at adapting to flaws in loudspeakers and find pleasure in music and movies in spite of them. But, why should we have to?

My very first simple blind experiments showed that non-auditory factors are powerful biases. Subsequent tests have proved it. We now know that, given the opportunity to judge sound quality alone, humans *with normal hearing* yield remarkably repeatable subjective ratings on loudspeakers, providing the key to understanding what measurements mean. Experienced listeners, with or without specific training, agree on what sounds "good". They are so reliable that they are worthy of the title "measurements". From published statistics on hearing loss, it seems that approximately 75% of the population qualify for this purpose as having sufficiently "normal" hearing: within about 20 dB of audiometric zero over most of the tested frequencies. So, we are looking to satisfy the needs of most listeners, not an exclusive few.

Sadly, those with significant hearing loss, whatever the cause, are less consistent in their judgments. They simply do not hear all of the sound, and therefore give different opinions at different times when listening to the same sounds. The perceptual handicap can go beyond sound quality, affecting one's perceptions of direction and space, something related to having difficulty conversing in restaurants and other acoustically complex venues.

I was a perceptive and reliable listener for decades but realized in my mid-60s that I was having more difficulty making judgments, and increased variability in my ratings showed that it was time to retire from the listening panel. Age and accumulated noise exposure had taken their toll. Hearing is precious, to be preserved. Those of us in this situation still enjoy music and movies, but our opinions are our own, distinctive, not for sharing.

Figure 1.1 conveys an important message: meaningful technical measurements are the goal, but reliable, repeatable subjective ratings of sound quality are essential to reach that goal. Normally one gets opinions from sighted evaluations, and these are demonstrably unreliable (Toole and Olive, 1994). Getting trustworthy subjective data is time-consuming and requires preparation and discipline. The technical side of the investigation is measurements, which many people can do, but they need to be the right measurements.

Measurements made in an anechoic chamber, or their signal-processed equivalents are the best descriptions of sounds that radiate from loudspeakers. Most loudspeaker listening is done in small rooms, with strong early reflections, so these data need to be processed and displayed in a manner that relates to what listeners hear. The fact that most rooms are at least somewhat rectangular, with vertical walls and horizontal floor and ceilings simplifies matters because those surfaces are responsible for the strongest early reflections—the second loudest sounds a listener hears. Measurements in anechoic chambers or the equivalent can yield masses of data, but not all of it is important to perceptions of sound and spatial qualities in rooms. Subjective measurements are the key to identifying those factors that influence our perceptions. In pursuing this rigid methodology, the hope was that it would lead to measurable performance targets, allowing engineers to design loudspeakers with confidence in how they would sound.

The rest of this book is a description of my near 60-year involvement with a research project that included others working with me, and recognizing relevant research done by others around the world. The goal is to provide a comprehensive scientific background to sound reproduction.

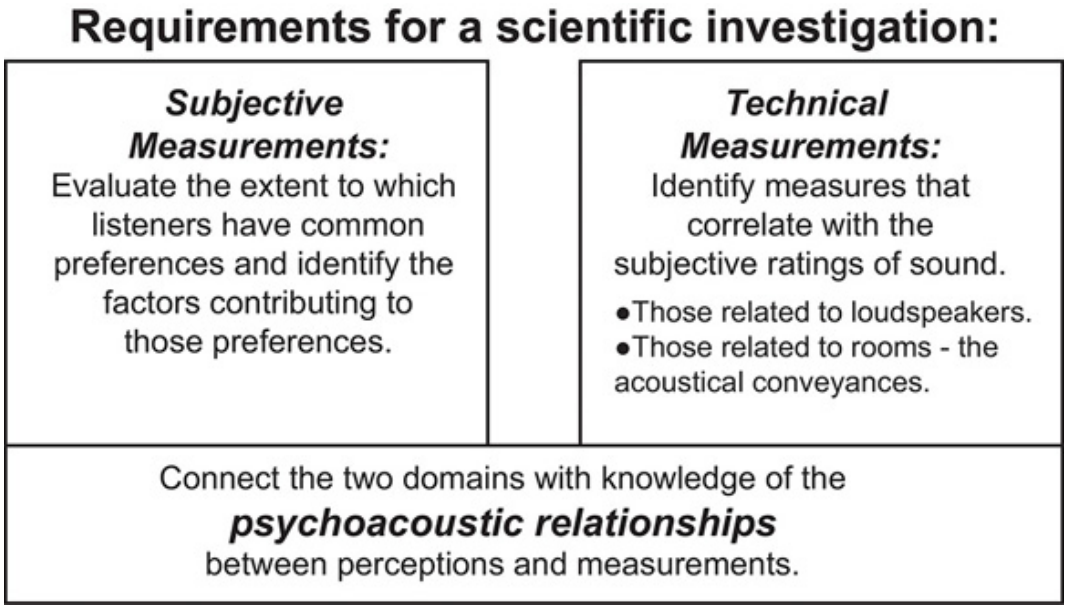


FIGURE 1.1 *The elements of scientific investigations in audio.*

This research began with me, and you will be reading much from me. So, let us begin by understanding who I am, and how I happened to end up dedicating my career to investigating “Sound Reproduction”.

1.2 FORTUITOUS EVENTS LEAD TO A CAREER

I was born in 1938 and grew up near Moncton, New Brunswick, Canada. My memory of reproduced sound from my early years was that table radios sounded “small”, lacking bass, and rendering everything with easily audible colorations of various kinds. Large console radios in polished wood cabinets sounded “fuller” but also sometimes duller. In those days people sought “good tone”, and sounding “mellow” was a compliment.

I took an interest in audio as a teenager in the era of fading 78s, the introduction of LPs, and, more exciting, stereo itself in the late 1950s. “Hi-fi” loudspeakers were large boxes, ugly and expensive, so doing it yourself was popular. Building magazine-design loudspeaker enclosures with my father was involving, as was the assembly of electronics using surplus materials from WW2 and kits. Every new do-it-yourself loudspeaker system sounded different, and the differences were not subtle. It was assumed that we all “hear differently” and had different tastes in sound quality.

None of those loudspeakers had the bass extension we are accustomed to now, and sensitivity—playing louder—was an important criterion. Those were the days of vacuum tube power amplifiers of limited power output. I built a version of the popular Williamson power amplifier, employing a premium Acrosound transformer, push-pull 6L6s with about 24 watts output. Most amplifiers of the period were under 15 watts. With such limited power, loudspeaker sensitivity was an important parameter, and horn loudspeaker designs were common. In those days the enclosures and loudspeaker drivers were separate items, rarely designed together, and many used off-the-shelf crossover networks for two-way systems. Knowing what we know now, they could not have worked well, and most didn’t. It was a time of trial and error.

Electrical engineering was my chosen field of study at the University of New Brunswick, where I received a BSc EE in 1960 and was awarded a Lord Beaverbrook Overseas Scholarship that took me, all expenses paid with a living allowance, to Imperial College, in London, England, for 2 years. There I intended to specialize in solid-state circuit design, capitalizing on the recent invention of the transistor. When I registered, I was stunned to hear that, due to high demand for the course, students from "the colonies" were denied access, even though I had a letter of acceptance. I was 22, alone in a foreign country, faced with finding something else to study, probably changing my intended profession, or go home. It was a test of my resolve.

After a few false starts in finding a motivating PhD project, I encountered recently retired Prof. Colin Cherry, and a conversation developed. Dr. Cherry was highly regarded in the field of human communication, including aspects of binaural hearing, notably the "cocktail-party effect": how it is that we can focus attention on one voice among many (Cherry, 1980). He also had thoughts about how stereo localization functioned. Several discussions led to a proposed doctoral research project based on exploring binaural hearing. At the time, most such investigations were conducted by researchers using limited signal generation and measurement schemes. It was thought that using my ability in circuit design and signal analysis to generate specialized sounds could lead to creative examinations of the hearing process.

Dr. Bruce Sayers, one of Dr. Cherry's last PhD students, then on staff at Imperial College, became my supervisor and the project was launched. The investigation used headphones, so that sounds to the two ears could be independently manipulated. Sounds were localized within the head, not externalized, but this was regarded as the beginning of a sequence of elaborations leading ultimately to listening in normally reflective rooms. The headphones used were highly regarded at the time, Sharpe HA10s made in Canada. I discovered later that they were designed by Dr. Edgar Shaw, who occupied the adjacent office when I joined the National Research Council of Canada after graduation. They employed the hard

shell, liquid-filled cushion around-the-ear design that is still employed in hearing protectors, co-invented by Dr. Shaw. It is a small world.

The research was productive, challenging the prevailing theories of hearing, and leading to three peer-reviewed papers in the *Journal of the Acoustical Society of America* by the time I graduated (Sayers and Toole, 1964; Toole and Sayers, 1965a, 1965b). The work is still quoted in textbooks (e.g. Blauert, 1996), which is gratifying.

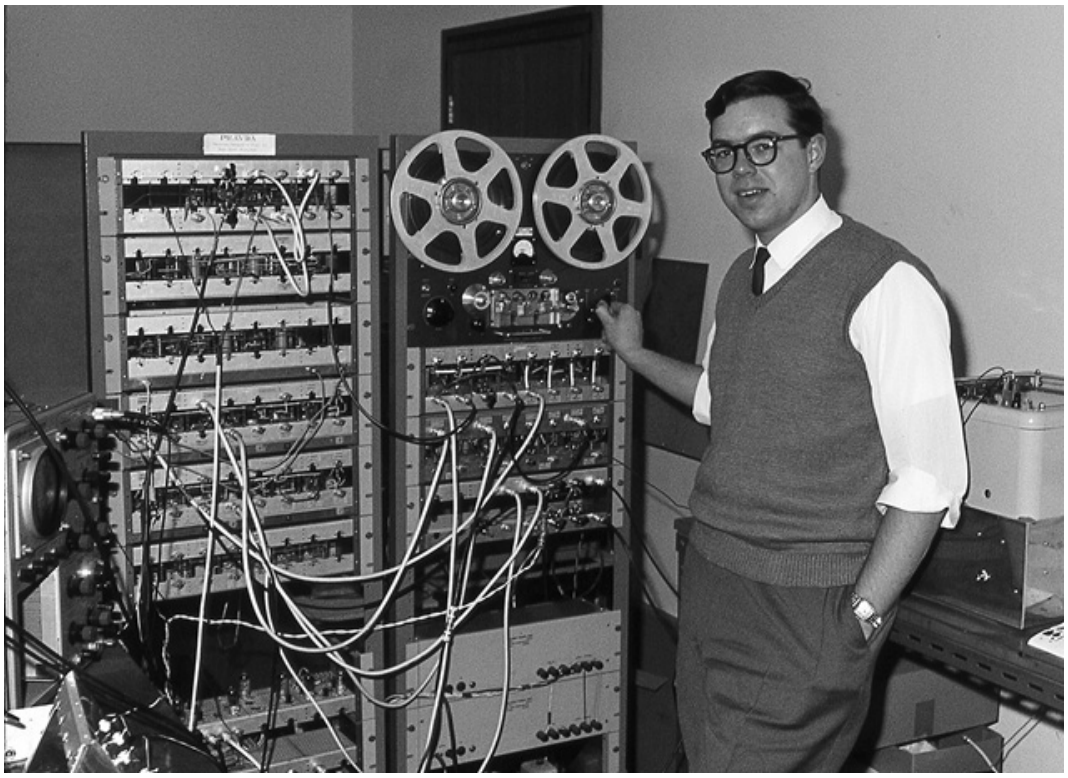


FIGURE 1.2 *Psychoacoustics research, engineering style. Racks of Toole designed and fabricated germanium transistor and tube electronics to generate and manipulate binaural signals, including a four-track pulse-width-modulated (DC—200 Hz) analog tape system designed to control randomized experimental parameters, and to store listener responses. Results were printed out on an automated analog X/Y plotter. Imperial College, London, ca. 1963.*

The last 2 years of study were supported by a Special Scholarship from the National Research Council of Canada (NRCC), a highly regarded, federal government–funded research establishment with activities in all major branches of the physical sciences and engineering. This led to an interview and a job offer as a research scientist in the Acoustics Group of the Applied Physics Division. It was a good fit for both parties, and in 1965 my wife and I moved to Ottawa, Canada.

1.3 MORE CHANCE FACTORS STEER THE RESEARCH

The Acoustics Group of National Research Council of Canada had an anechoic chamber with 1 m wedges, which was a requirement for the next phase of my sound localization investigations—a reflection-free space. Technically accurate loudspeakers were needed to deliver the contrived test sounds, so I measured the performance of the loudspeakers that were present in the lab, an Acoustic Research AR-3 and a Tannoy Cadet. What was found was disconcerting. The measured anechoic on-axis frequency responses were far from flat and very different from one another. At the time the AR3 was a highly respected loudspeaker, so there was the question of how such a non-flat performance could sound good. All electronics, then and now, aimed for a flat frequency response over the audible frequency range so as not to modify the recorded sounds. Why should loudspeakers be different?

Curiosity required further investigation, so I borrowed other hi-fi loudspeakers from colleagues and measured them. They too were not flat, and all were different. Up to this point, I had only seen “specifications” for frequency response and, if it were not for the unimpeachable pedigree of the measurement circumstances, it would have been possible to think that there had been a tragic error in making the measurements. Suddenly, claims that the loudspeaker was the “weakest link” in the audio chain rang true. But could these products really sound as bad as some of the curves looked?

A “Friday-afternoon” experiment was arranged in late 1965 to do a simple comparison listening test in one of the laboratory rooms. Having

learned the basics of experimental psychology for the doctoral thesis work, it was obvious that this test had to be somewhat controlled. So, cotton sheeting was hung to render the experiment “blind”. The loudspeakers were compared in monophonic A/B/C/D randomized sequence and were adjusted to be equally loud. There was no special reason for listening in groups of four loudspeakers; it was just expedient. A supportive technician built a simple level-matching relay switch box. After that, several people took turns listening to a variety of music, forming opinions, and making notes. A “gestalt” impression, a summarized overall sound quality/preference rating, was required: a number on a scale of 0 to 10, where 0 was rubbish and 10 was perfection. No discussion was permitted during the tests.

The results surprised everybody. The audible differences were enormous, but there was general agreement about which ones were judged to sound good. It remained a topic of discussion for days. Where were the expected personal preferences? And, equally interesting, was the observation that the most preferred loudspeakers had the best-looking anechoic measurement curves, suggesting that flat and smooth might be the objective. Some of the loudspeakers did, in fact, sound as bad as their measurements looked.

The need for loudspeakers for my sound localization experiments remained, and the winner of this simple test, a KEF Concord, showed promise. I had purchased a pair of these loudspeakers for my post-graduation hi-fi system, based on listening in tiny rooms at the 1964 audio show in the Hotel Russell in London. To my uneducated ears it simply sounded good. Raymond Cooke, owner of KEF, gave a discounted price because it would be for “research”. That was a fib at the time, but it turned out to be true.

As good as it was, there was some midrange misbehavior which was identified to be the result of resonant flexure along the long axis of the racetrack shaped B139 woofer. An absorbing phase plug over the center 1/3 of the diaphragm eliminated most of the cancellation and a new crossover resulted in an improved loudspeaker. My first loudspeaker “design”. The sound localization experiments proceeded.

1.3.1 The Listening Test That Changed My Career

Months passed before another listening test was staged in February 1966. By then I had learned that bed sheeting is not acoustically transparent and that music passages should be short and repeated so that more meaningful comparisons can be made. Every musical passage needed to be heard through every loudspeaker. This meant that a "disc jockey" was needed to perform the tedious task; it was the LP era. Word had spread and audio enthusiasts from within the organization lined up to participate, in some cases bringing their personal loudspeakers to be evaluated. Again, the multiple comparison method of listening was employed, with the four selections from a population of six loudspeakers being randomly changed. This test went on for a few days and yielded enough subjective data to warrant rudimentary statistical analysis. There was an overall pattern of preference in the global "gestalt" sound quality ratings, but it was surprising to note that the two musicians in the listener population were quite inconsistent, at times liking the products rated lowest by the majority. It has subsequently been shown that this is a real phenomenon. Musicians who are not also audio enthusiasts may make judgments on very different rationales, including "musically valid" interpretations of instrument sounds. Their perspectives on instrumental sounds are different from those of concert attendees. Unfortunately, they may also suffer from hearing loss, which is an occupational hazard. It was an indicator of a possibly significant variable.

In these tests there was again good agreement about the products that were preferred and those that were not. The highest rated loudspeaker was my redesigned KEF Concord that was being used in the anechoic chamber binaural investigations, which had been redesigned for "technical accuracy". Figure 1.3(e) and (f) show that loudspeaker before and after my modifications. It had the best-looking set of measured data, confirming the perspective that there is value in smooth and flat frequency responses on and off axis. In subsequent subjective evaluations, this loudspeaker remained "king of the hill" for an extended period. The original handwritten response sheets from these tests done 58 years ago still exist.

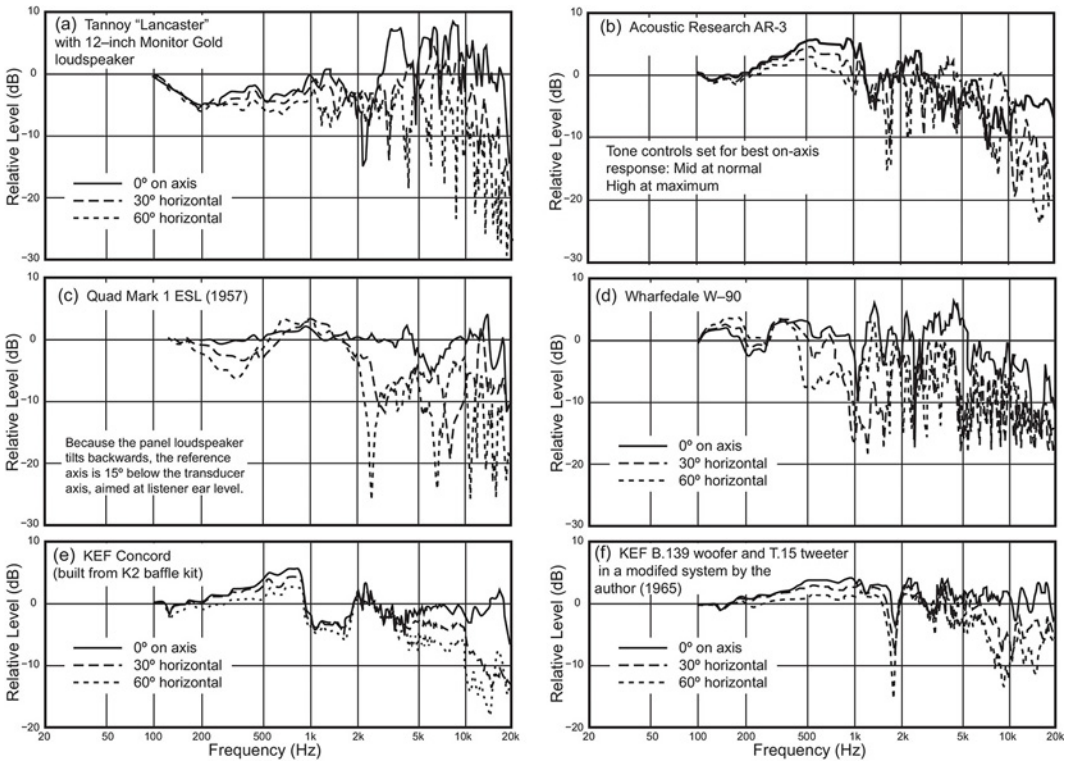


FIGURE 1.3 *On- and off-axis anechoic measurements on five loudspeakers that were highly respected in the mid-1960s. The sixth (f) was (e) as modified by the author. Low-frequency data are truncated because the anechoic chamber had not yet been calibrated.* ⏎

Figure 1.3 shows anechoic on-axis and 30° and 60° off-axis frequency responses of six loudspeakers that were evaluated subjectively. The curves are all different, and all exhibit serious deviations from the theoretical ideal horizontal flat line. It seems that some of the designers attempted to be flat on axis but did not succeed. The highest scoring loudspeaker, the redesigned KEF system, had the most orderly and flattish curves.

The original Quad ESL (Figure 1.3(c)) had a good on axis behavior, but the large radiating panels degraded off-axis performance, causing in-room performance to be compromised. The "dual-concentric" Tannoy (a) was an attractive concept, but it radiated annoyingly bright and colored high frequencies along with very boomy bass. The multi-driver large box Wharfedale (d) exhibited numerous resonances, and abundant acoustical

interference. It was very colored and lacked sufficient low and high frequencies.

The highly regarded Acoustic Research AR-3, shown in Figure 1.3(b), embodied Edgar Villchur's innovative acoustic-suspension woofer, offering deeper bass in a small enclosure than had previously been possible. The dome midrange and tweeter units had excellent dispersion but being surrounded by a deep grille-cloth frame generated diffraction and acoustical interference "grass" in the curves. AR had released an improved AR-3a, so I took the AR-3 to have it upgraded at the source, and to have a tour of their facility. Subsequent re-evaluations were similarly disappointing, because the basic spectral shapes remained. Much about the product was well executed. The underlying problem was a deliberately planned, but inappropriate, frequency response target, as discussed in detail in chapter 18.1 in the third edition of this book (replicated in OR Chapter 1).

These findings were thought-provoking. The loudspeaker industry was large and worldwide, and these data showed little evidence of standards guiding it. That people found all these loudspeakers attractive enough to purchase underlined the importance of blind listening if the objective is to evaluate sound quality. The influences of non-auditory factors like brand, marketing claims, size, appearance, price, etc. are obviously significant. There was little evidence of the "we all hear differently" school of thought, but implications that over time we can adapt to find pleasure from significantly flawed loudspeakers. Most importantly, the results showed that there was a clear relationship between what was measured and what was heard, and that something approximating flat and smooth, technical accuracy, was the subjectively preferred target anechoic on-axis frequency response. It made sense. Off-axis behavior would be something to explore, as we listen in reflective rooms.

The director of the Applied Physics NRCC division within which I worked had participated in the blind listening evaluations, and to his chagrin, he disliked his own large, expensive, widely respected Tannoy loudspeakers. This may have influenced the subsequent discussions of whether this was a possible line of future research. It was approved as a

topic for preliminary investigation. Over the next few years, I gradually reduced activity in the binaural hearing research, and the loudspeaker work became a full-time project—it was a "here and now" problem for a global industry, for which there seemed to be a solution.

1.4 SUBJECTIVE EVALUATIONS MATTER, AND HOW THEY ARE DONE MATTERS MORE

Pondering the implications of the preliminary test results, it is evident that the subjective evaluations, the opinions, were not random, or strongly personal, and they appeared to be related to prominent features in the anechoic measurements.

Acoustical measurements can be done in many places, and if done properly, the same curves will result. Acoustic Research had published anechoic chamber data on their AR-3 loudspeaker, and I was able to closely duplicate their measurements. Their product got low ratings because they were aiming at the wrong performance target, not because they could not measure their product. They later changed their target and went on to produce some excellent neutral sounding loudspeakers. Not every manufacturer has an anechoic chamber, but anyone can measure outdoors, weather permitting (as AR did for some of its data), and nowadays there are elegant digital methods that generate the same data in reflective measuring venues. Some manufacturers used steady-state in-room measurements, which, as will be shown, are neither usefully analytical nor definitive of sound quality. The research challenge is to determine which measurements to make and how to interpret them.

Subjective evaluations can provide guidance, but they must be done in a fashion that yields trustworthy results. My first paper to the Audio Engineering Society was titled "Listening Tests—Turning Opinion into Fact" (Toole, 1982), and it summarized what I knew about the topic at the time. The methods were sufficiently good that I began to call the results "subjective measurements", as shown in Figure 1.1.

Every manufacturer evaluates loudspeakers for their potential appeal to customers. The results determine what product will be manufactured for sale, duplicated hundreds or thousands of times. Often these evaluations are informal affairs, almost always sighted, probably in stereo, probably including company officials. Stress levels are high, reputations and egos are challenged, and the decision makers may or may not be good listeners. I have witnessed a few of these uncontrolled sessions, and in one I recall seeing the expression on an engineer's face when the brand manager, after listening to a few bars of music through his well-designed loudspeaker, pronounced "more bass, more treble", and left. Often the choices are biased by the marketing philosophy of replicating what sold well last year, even a competitor's product, not having confidence that something different might be better. The wide variety of sounds in the marketplace over the years had their origins in substantially uncontrolled subjective evaluations, not design engineers aiming at a particular version of non-flat frequency response or a specific collection of mechanical and acoustical resonances.

In loudspeaker evaluations, listeners are presumed to be able to separate the sound signature of the loudspeakers from those of the room and the music programs. That is not easy.

All evaluations yield opinions, but not all opinions are answers to the question being asked.

- The first, most basic requirement is that they be done blind, so that non-auditory factors are eliminated. Obviously price, appearance and size are important factors in the final choice for consumers but, when evaluating *sound quality*, blind listening is necessary.
- The second requirement is that when making comparisons the loudness levels be equalized, so small differences in sensitivity don't influence the sound quality ratings.
- How the listening presentations are arranged influences the ease with which listeners are able to isolate the sound contributions of the

loudspeakers, as separate from those of the room and those inherent in the programs.

- Finally, it matters whether the evaluations are done in mono or stereo. Mine were done in mono because it was simple, and it seemed that we were evaluating the inherent performance of the loudspeakers, not how they behaved in stereo. But that would be tested.

Let us review the popular options.

Single Stimulus Method. A common method of evaluating sound quality involves listening to a single loudspeaker type, in mono or stereo—"*single stimulus*". In it the loudspeaker and room sounds are combined and constant during the evaluation; the program is variable. Such reviews might be better described as record reviews, not loudspeaker evaluations. In the absence of comparison loudspeakers, adaptation sets in, and "familiarity" may be the deciding factor in the opinion being formed. Listeners must rely on long-term acoustic memory of how the room and various recordings sound, and this is known to be capricious. This is the widely practiced "take it home and listen to it" method. Notable contradictions abound in the resulting opinions, partly because there are no comparison sounds available, and auditory memory is short. A significant factor is that bass accounts for about 30% of our overall impression of sound quality. It is dominated by room resonances and all listening room setups are distinctive, requiring measurement and management (a separate chapter) if results are to be comparable. Finally, "breaking in" is sometimes mentioned.

Breaking in is the belief that all audio products, electronics, wires and loudspeakers, become more acceptable, more musically proficient, after a period of use (tellingly, it is never the reverse). In spite of opinionated essays from reviewers, many of which echo marketing claims, I am not aware of any scientific evidence of audible performance change in electronics, wires and loudspeakers, except for a minor change in low-bass performance of woofers and subwoofers

as the mechanical suspension elements become more compliant during early use. This will have a small effect on the lowest bass frequencies, which may or may not be noticed by listeners. Such differences are difficult to discern among the distractions of strong resonances in small rooms, and the varying bass content in recordings. In general, engineers design products *not* to change with time—wine does that. Otherwise, audio products would have vintages. What is happening is that listeners are adapting to a new timbral signature, or just imagining things. The human brain is good at that. Manufacturers that promote break-in at anything other than low bass frequencies of loudspeakers may just be catering to the folklore, or they may have products that actually need some time to get used to—a caution.

Paired Comparison Method. The classic A vs. B *paired comparisons* test method exposes audible differences, which is good, but may not reveal problems shared by both test objects. Subjectively, it allows listeners to begin to separate the sound of the loudspeaker from those of the room and the program because both are constant factors in the comparison. To gain perspective on where the two test objects stand in the hierarchy of loudspeakers in general, multiple paired comparisons including other loudspeakers randomly selected from a population need to be staged. This adds complications and takes more time but is amenable to statistical analysis. This is important, as it gives confidence that the resulting ratings are not merely due to uncontrolled variables. But even statistically significant differences can be misleading information in some circumstances. For example, if the same loudspeaker is always used as a standard reference, a flaw in the reference can become a sought-after virtue. Some reviewers and manufacturers have fallen into that trap. Adding loudspeakers with different designs and transducer configurations to the comparison population aids in revealing problems, which leads us to the next method.

Multiple Loudspeaker Method. In my first crude tests, listeners easily identified the distinctive characteristics of the test loudspeakers. The *multiple loudspeaker* method was used because it was simple and quick. In retrospect, it was a fortunate choice. Comparing multiple loudspeakers at one sitting in the same room gives listeners the opportunity to perceptually separate the distinctive timbral contributions of the loudspeakers from those of the program material, a constant factor, and the room, which will have some constant factors and others, primarily at low frequencies, that are dependent on loudspeaker and listener locations.

In any comparison test, if loudspeakers being compared are in different locations, additional tests are required with the locations randomized to average out the room effects. This necessity adds time and tedium to the test, so it is not always done. If listening tests are to be a much-repeated activity, this alone is motivation to create an apparatus to substitute different loudspeakers at the same location in the room, so that acoustical coupling to low-frequency room resonances and adjacent boundary interactions are constant factors. The improvements in test efficiency and effectiveness are profound—important to manufacturers and reviewers. Options for how to do positional substitution will be discussed later.

ABX Method. The well-known ABX method tests whether listeners can identify which of A or B is the same as X. It is designed to determine if listeners are able to hear a difference—any difference, however small—not to scale the magnitude of the difference. Such tests often lead to black/white, winner/loser interpretations, when in a realistic context, they may both be winners, but only slightly different. The method and the apparatus are regularly criticized when they do not reveal an audible difference between products having different marketing stories. The truth sometimes hurts.

MUSHRA (ITU-R BS.1534–3) and ITU-R BS.1116–3 Methods. These are specialized tests for evaluating the significance of very small errors in codecs. The introduction to BS.1116 states: “This Recommendation is intended for use in the assessment of systems which introduce impairments so small as to be undetectable without rigorous control of the experimental

conditions and appropriate statistical analysis. If used for systems that introduce relatively large and easily detectable impairments, it leads to excessive expenditure of time and effort and may also lead to less reliable results than a simpler test." These test methods are not relevant to loudspeaker evaluations, but the loudspeaker choice and calibration method, and acoustical recommendations of BS.1116 are sometimes referred to, and they need careful consideration before use (see OR Chapter 12).

Subjective Rating Scales Are Non-linear. Helson's (1964) adaptation-level theory addressed the lack of stability in subjective rating scales. Listeners are hesitant about using the extreme values of a scale, probably leaving space in the event something better or worse comes along. It is also clear that special efforts are necessary if anything more than a "better/worse" result is desired. The solution was to provide listeners with "anchor" products, randomly inserted into experiments to remind listeners what good and bad sounds were really like. We kept a small stock of loudspeakers with very high and very low subjective ratings for this purpose. It cannot be perfect, but it helped.

Figure 1.4 is my conceptual representation of the elasticity of subjective ratings of loudspeaker sound quality, and how listeners tend to "stretch" the subjective rating scale to fit a comfortable range in response to context changes. Products that are excellent, and closely competitive, relative to other products representing a large population, get lower ratings as poorer products are removed from the tests. Listeners still identify the rank order of sound quality among the products but exaggerate the magnitude of the differences in the absence of low-rating "anchor" products reminding them of how bad sounds can really be. Subjective rating scales are elastic. Fortunately, winners remain winners. However, design engineers and consumers both also consider cost, and that is why some degree of discipline in stabilizing the rating scale is important—a slightly lower-rated but much less expensive product may be advantageous from both sales and personal purchasing perspectives.

Zielinski et al. (2008) provides an exhaustive academic examination of biases in subjective ratings. It is enough to deter all but the most determined from even attempting such an exercise. Fortunately, in 1965–1966 none of this was known. Visual inspection of scatter diagrams turned out to be sufficiently useful guidance to justify the research and to identify the important experimental factors that allowed it to continue. Humans are good at recognizing trends in data distributions—statistical validation came much later.

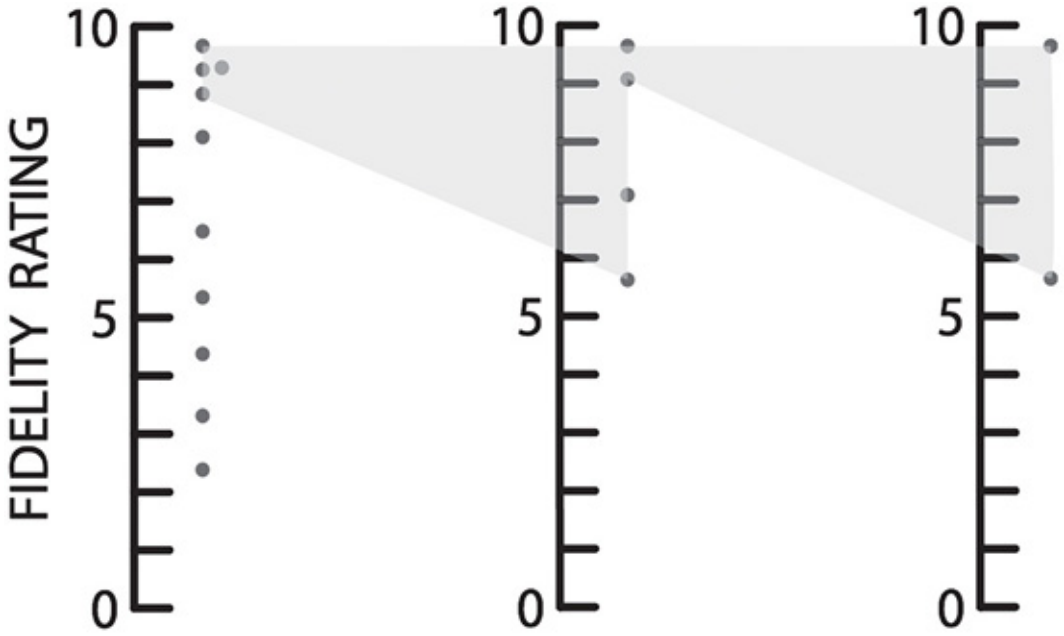


FIGURE 1.4 *A conceptual representation, not an experimental result. On the left, listeners evaluated ten products in multiple randomized tests, identifying four closely rated very good ones. In the middle, the four closely rated top products were evaluated in a multiple comparison test; the listeners used more of the subjective rating scale in the absence of poorer products. Finally, on the right, two of the top-rated products are compared in an A vs. B test, and listeners again expanded the rating scale, exaggerating the real difference between these products. From Toole (1998).* ⏎

Opinions differ on the merits of experimental methodologies, and the multiple comparison method was little known. Schuck et al. (1993) undertook an elaborate comparison of multiple vs. paired comparison listening methods, including comparisons of different statistical analysis models. They found that the same conclusions would be reached using any of the methods, but noted:

> A disadvantage of the paired-comparison paradigm in the case of a 4 loudspeaker test, is that 6 times as many trials are needed to acquire the equivalent information that can be acquired in using a multiple-comparison paradigm; when the experiment requires the experimenter to move loudspeakers at the beginning of every trial, this becomes a major disadvantage.

They added:

> When actual ratings of the Fidelity Scale are sought from the subjects the multiple-comparison paradigm apparently results in a wider spread of ratings with larger standard deviations than the paired comparison paradigm. However, the power of the method to detect loudspeaker differences is greater for the multiple-comparison paradigm.

Auditory memory is always an issue. How long can we accurately remember a timbral identity for comparison to another? Pike et al. (2014) investigated the effect of auditory memory on loudspeaker ratings made after different quiet intervals between auditory samples. The music samples they used were repeated, beginning and ending before switching, so the listener is comparing the end of the sample with the beginning of the sample after an interval. These are different sounds and audible contrasts are inevitable. They said that:

> Regardless of the exact mechanism, the data suggest that a period of silence of approximately 500 ms–1 s is beneficial to discrimination accuracy. *It should be noted that the situation in this test is different from that in some standard audio product listening tests where the listener can switch immediately between differently processed versions of a program item* [my italics] without restarting playback. With this sort of continuous playback switching, performance may be better than is suggested by the 100 ms RI *(retention interval)* condition. This is likely

to be because only the applied processing is changing rather than there being an additional change in spectrum between the offset of one stimulus and the onset of another.

Therefore, the practice of having instantaneous switching with running, repeated (looped) program excerpts should be an advantageous test methodology. The pleasure in the program is lost, but the evaluations are more revealing. This is the method used in the evaluations reported in this book.

In the present investigations, the magnitudes of the differences are important; consequently, efficient three-way and four-way multiple comparisons have remained the custom in our subjective evaluations. The fact that the comparisons were done using single loudspeakers, in mono, when we commonly listen in stereo, has been a source of controversy. But, as will be discussed in Sections 4.3 and 4.4, stereo has its own problems. Good loudspeakers are appreciated in all listening formats.

1.5 A REVIEW OF SOME MEASUREMENT OPTIONS

First, we want to identify measurements that reveal audible differences. Then, we need to discover what distinguishes those measurements that correlate with high subjective ratings of sound quality. The goal is to give design engineers the ability to recognize potentially good-sounding loudspeakers by examining the right set of measurements, presented to the eyes in an optimum fashion.

Performance measurements of audio devices fall into two general categories: linear and non-linear behavior. Linear characteristics are those that do not change with sound/signal level: magnitude vs. frequency (a.k.a. frequency response) and phase vs. frequency (a.k.a. phase response). These two can be mathematically transformed into the time-domain impulse response (a.k.a. transient response). The data are interchangeable.

In “minimum-phase” systems and devices the time-domain behavior, the transient response, is predictable from the frequency response alone, which may seem like magic but is real. Fortunately for those of us concerned with

sound reproduction, loudspeaker transducers within their operating ranges (not the crossover regions of multi-way systems) and prominent low-frequency resonances in small rooms behave as minimum-phase systems. This means that simple parametric equalizers can be used to attenuate resonances in transducers at any frequency and conspicuous resonances in rooms at low frequencies, with the result that the time-domain misbehavior (ringing) is reduced when resonant peaks are reduced. Flat and smooth frequency response is the non-resonant ideal. These are huge advantages when loudspeakers and electronics are integrated with or dedicated to each other. The anechoic performance of the loudspeaker can be optimized, with resonances electronically attenuated.

1.5.1 Resonances Are Important

Almost all of the sounds we are interested in are composed of collections of resonances. Voices and musical instrument sounds are combinations of mechanical resonances (vibrating vocal cords, strings and wooden panels connected to them, vibrating reeds or lips in a horn mouthpiece, skins of drums, etc.) and acoustical resonances (mouth and nasal cavities, resonances in the complex cavities within stringed instruments, resonances in the tubes of brasses and woodwinds, pipes of organs, etc.). To these are added scraping, clicking and aerodynamic/breathing sounds, all of which add up to distinctive, highly recognizable timbres of the sound generated by people and instruments. If loudspeakers add resonances when reproducing such sounds, they modify the timbre, which listeners might recognize. The problem is especially annoying because resonances in loudspeakers modify the timbre of *all* sounds played through them in the *same* way, adding a repetitious component to what is heard. This is why they are easily detected in comparison listening tests. In real life, we can adapt to a certain amount of this. We all have enjoyed music and movies through inferior loudspeakers, but when presented with a comparison to more neutral reproduction, the unwanted resonances stand out and are judged negatively. In other words, we can live with less than perfect sound, but if given the choice, we prefer neutral, non-resonant reproduction. In terms of

identifying useful loudspeaker measurements, those that permit the recognition and evaluation of resonances are the most valuable. *A single curve is not sufficient*. This will be explained.

Mechanical and acoustical resonances can arise in the transducers, and in the enclosures of loudspeaker systems. Much ado has been made about mechanical resonances in loudspeaker enclosures, for good reason. They are large surfaces and can radiate significant sound if they move. When loudspeakers are measured in an anechoic chamber, or equivalent, and examined through 360° of rotation, any such radiated sound is captured, and, if significant, its presence will be noted. However, accelerometer measurements of the mechanical movement at a point on an enclosure surface is not definitive evidence of effectively radiated sound energy, or of its audibility. Some resonant modes in panels exhibit similar areas simultaneously moving in opposite directions. There is movement, but little sound is radiated. With scanning laser interferometers, it is possible to anticipate the sound radiating modes and provide selective stiffening or damping to address the problems. Brute force methods, like massively thick, heavy materials also work, of course, but elegant engineering can be effective and keeps cost and weight down.

On one of my visits to Japan, my attention was drawn to a popular brand that had an inch of concrete poured into the top of every loudspeaker. This was so that it would pass the “knuckle test” exercised by shoppers as they moved around the store floor tapping the loudspeakers. That was clever, but the knuckle test itself is not. It excites the panel at the point of contact, not as it is by sound inside the cabinet, which is pressure distributed over the surface. A well-designed loudspeaker might not pass the knuckle test. More folklore.

1.5.2 Non-linear Distortions

Non-linear distortions are additional frequencies generated when signals or programs interact with a device that changes its behavior as a function of sound level—the moment-by-moment input-to-output relationship is not a straight line. The non-linearity might be in, for example, a woofer cone

suspension that at a point requires proportionally greater effort to move as the cone displacement approaches it limit. It may sound clean at low sound levels, but as the volume is turned up, the timbre changes as additional sounds (distortion products) are generated by the non-linearity. There are also other causes.

The problems originate as measurable input/output non-linearity in devices, but the industry has chosen to evaluate the magnitude of the non-linearity by probing them with well-defined test signals and measuring the amount of added distortion products at various sound/signal levels. The simplest test signal is a pure tone, and the distortion products of pure tones are harmonically related overtones; therefore, we have "harmonic distortion". Music is not simple pure tones, and the broad spectrum of music that generates the distortion products also makes them less audible—they are perceptually masked. As a result, harmonic distortion, as a measure, is not a reliable indicator of its audibility. To design engineers, it is a useful *relative* measure when designing systems within their context, but as an indicator of potentially audible annoyances for consumers it is almost useless. The only truly safe value for harmonic distortion is zero, or very close to it.

Other, more complex methods of quantifying non-linear behavior exist. Intermodulation and multitone methods definitely add information. These probe the non-linear device in different ways, generating distortion products having different spectra. However, all numbers generated are ultimately of questionable value unless they incorporate some means of compensating for perceptual masking. The amount of perceptual masking varies with the spectrum, sound level and temporal nature of the program that creates the distortion, so it is not at all straightforward. Improvements are possible, but now it is still work in progress. We are a long way from having a standardized, trustworthy method of quantifying the audibility of non-linear distortions as they might be revealed in normal music and movies.

In the abundant small, inexpensive, limited bandwidth loudspeakers in the marketplace, non-linear distortions may be heard when they are pushed hard, although many of them have protective schemes in the electronics that

limit the sound output to avoid distortion. However, among mainstream hi-fi, audiophile, and studio monitor loudspeakers, audible distortion is rare if the loudspeakers are used within the intended limits. In the engineering of quality loudspeaker systems, power compression and non-linear behavior are routinely factored into the selection of drivers, crossover design, etc. If there are problems, they are likely only to occur when the systems are driven to their limits. At this point, one's ears may be non-linear and hearing may be at risk.

1.5.3 Frequency Response Is Important— But What Is It?

In strict engineering terms, the magnitude vs. frequency and phase vs. frequency responses together describe the frequency-domain performance of systems. They also describe the time-domain behavior of the system: the impulse response. The information in the frequency and time domains are interchangeable—one can be measured and converted into the other by calculation.

In general, the amplitude vs. frequency response dominates what we hear. It has come to be known simply as "frequency response" and it is a required performance specification. But, what about phase response, and phase shift in the sound delivery path?

Both amplitude and phase responses of transducers are essential engineering data when designing loudspeaker crossovers and arrays where the outputs from multiple drivers must acoustically sum in an orderly manner. However, once the sound has left the loudspeaker system, phase, by itself, turns out not to be a significant audible factor.

When put to the test, phase shifts, group delay and even complete polarity inversions turn out to be "is it there or not?" issues, requiring ABX blind testing. Using contrived signals in headphone or anechoic listening some listeners have reported hearing a "difference", but a reliably stated "preference" has been elusive. In normally reflective rooms using music the differences, if real, become vanishingly small or non-existent. Intuition fails again. More information will be provided in a later chapter.

Putting all this together, if amplitude response is a strongly audible factor, and phase response is not, it means that waveform details are relatively unimportant to auditory perceptions. Think about it.

However, both amplitude and phase response matter in the *matching* of stereo left and right loudspeakers that perform in double-mono to generate the phantom images in the soundstage. It is the similarity of the loudspeakers that is important, though, not the actual responses. Credible phantom images can be created by two flawed, but identical, loudspeakers.

In electronics and in headphones, a single curve is sufficient to describe the frequency response. However, the anechoic frequency response measurements of loudspeakers must include those sounds that are radiated off axis, as well as on axis. This is necessary to be able to describe the character of the reflected sound field in small rooms that the binaural hearing system responds to in different ways.

Consumers and professionals searching for loudspeakers capable of delivering neutral, uncolored, reproduced sound need informative measurement data, which must be both available and understandable. This book will explain the meaning of the necessary anechoic measurements, and the origin and meaning of a set of anechoic measurements called the spinorama, which describes the sounds as they may be heard in normally reflective listening rooms. This method is now part of an audio industry standard (ANSI/CTA-2034-A, 2015). At present, spinoramas are available from some manufacturers and through independent sources on the internet. Some manufacturers might prefer for such measurements on their products not to be available, and some marketing departments avoid publishing such data because it might not be understood, or, if understood could be embarrassing. The latter is ironic, because the most commonly published specifications, such as “Frequency Response: 32 Hz to 22 kHz □ 3 dB”, are useless as predictors of sound quality. It is a frustrating situation.

1.5.4 Room Equalization/Calibration Adds Confusion

If it is not possible to access spinorama or equivalent anechoic data to gain understanding of how a loudspeaker might sound, why not just measure,

and equalize/calibrate it in the listening room after installation? That way, the room acoustics are included, which sounds logical.

For decades it has been widely accepted that a steady-state amplitude response measured with an omnidirectional microphone at the listening location in a room is an important indicator of how an audio system will sound. These easily done measurements have come to be known as “room curves”. That belief has a long history in professional audio, including large venue and cinema sound, using pink noise as a stimulus and 1/3-octave-band analysis of the result. New versions of the method have infiltrated consumer audio with stand-alone products and receivers incorporating automated measurement and equalization capabilities. The implication is that by making in-room measurements and equalizing the input signal so that the room curve matches a predetermined target shape, imperfections in (unspecified) loudspeakers and (unspecified) rooms are measured and repaired. It is an enticing marketing story (Toole, 2015). The reality is that without detailed anechoic data on the loudspeakers it is not possible to interpret the audible implications of a steady-state room curve at middle and high frequencies. Bass performance, however, is correctly revealed.

It is implied that equalization in the electronic signal path can correct for aspects of the reflected sound fields in the listening room, avoiding untidy acoustical treatments. This stretches credibility. In-room measurements that display events in the time domain can be used to *identify* potential problem reflections, but these non-minimum-phase phenomena (the reflected sounds arrive from many directions at many times) can only be manipulated by sound absorbers or diffusers. It is essential to know *what* needs to be corrected, *how* it should be corrected, and what should simply be left alone. Figure 1.5 explains the problem.

1.6 LISTENING IN AND THROUGH ROOMS

Sound fields in rooms are complicated three-dimensional events, ever changing with the program content. One of the most complex sound fields we experience in life is that in a concert hall, with a stage that may contain

100 or more musicians. There are instruments of greatly varying directional properties—from highly directional brasses to widely, irregularly dispersing strings. The sound field in a concert hall is very diffuse, designed deliberately to blend and convey the music to all parts of the large audience. The direct sound, the first sound to arrive at your ears from the source, attenuates rapidly with distance and is quickly surpassed by the reflected, reverberant sounds. The point at which the direct and reflected sounds are equal is called the “critical distance”, and in many concert halls it is just a few rows into the audience. Yet, listeners well back in the hall can localize the approximate positions of instruments, especially those emitting transient sounds. Once localized, the perceived location tends to remain even when the continuing real-time sounds are confusing. This experience is likely associated with plausibility, in which it is not credible for a sound source we can see to be heard in different locations at different times. There is a limit, though: at very distant seats the localization becomes progressively blurred, but loudness is better sustained because it is determined by reflected sounds.

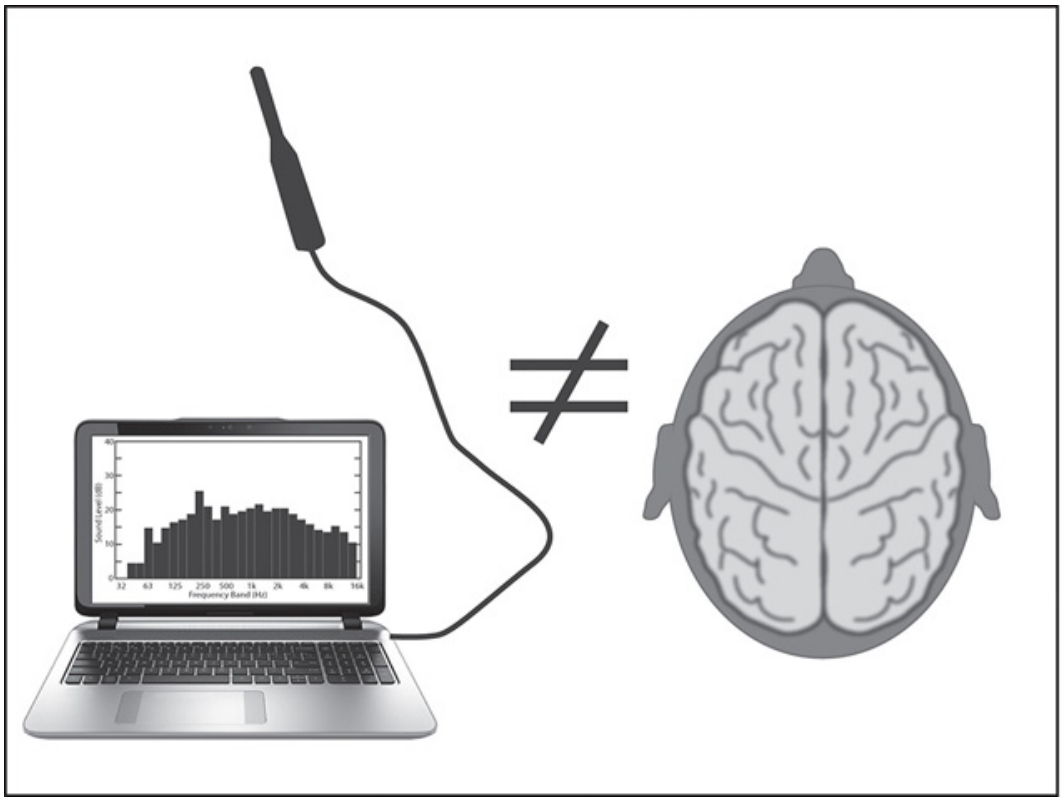


FIGURE 1.5 *What is measured in a room with an omnidirectional microphone and what is perceived by human listeners—two ears and a brain—are related but are not equivalent. In addition to timbral and spatial distinctions we make in perception, there is the fact that humans adapt to some aspects of room acoustics.* ⏎

Dr. Arthur Benade, a musician, physicist, and very insightful person proposed the center and left-side portion of Figure 1.6. I added the right column because it seemed appropriate.

The message in Figure 1.6 is that humans have some ability to "listen through" rooms. The sound field in any room, from concert hall to domestic listening room, is a complex mixture of direct sounds from the sources and countless reflections from all surfaces and objects within the room. At the listening position the direct sound from the sources arrives first, followed by numerous reflected sounds arriving from different directions at different times. Despite the complexity, the perception is normally of a sound source,

a voice or musical instrument, performing in an acoustic space. Change the space and the perception is of the same voice or musical instrument, but in a different space. The implication is that listeners appear to perceptually recognize acoustic spaces as separate entities. Bregman (1999) describes this kind of process as perceptual streaming, in which different portions of an auditory "scene" can be viewed somewhat independently—foreground and background, for example. If the audience is displeased with what is heard, as occasionally happens, the blame is placed on the hall, not the voices or instruments. We humans know the difference. Binaural hearing is powerful.

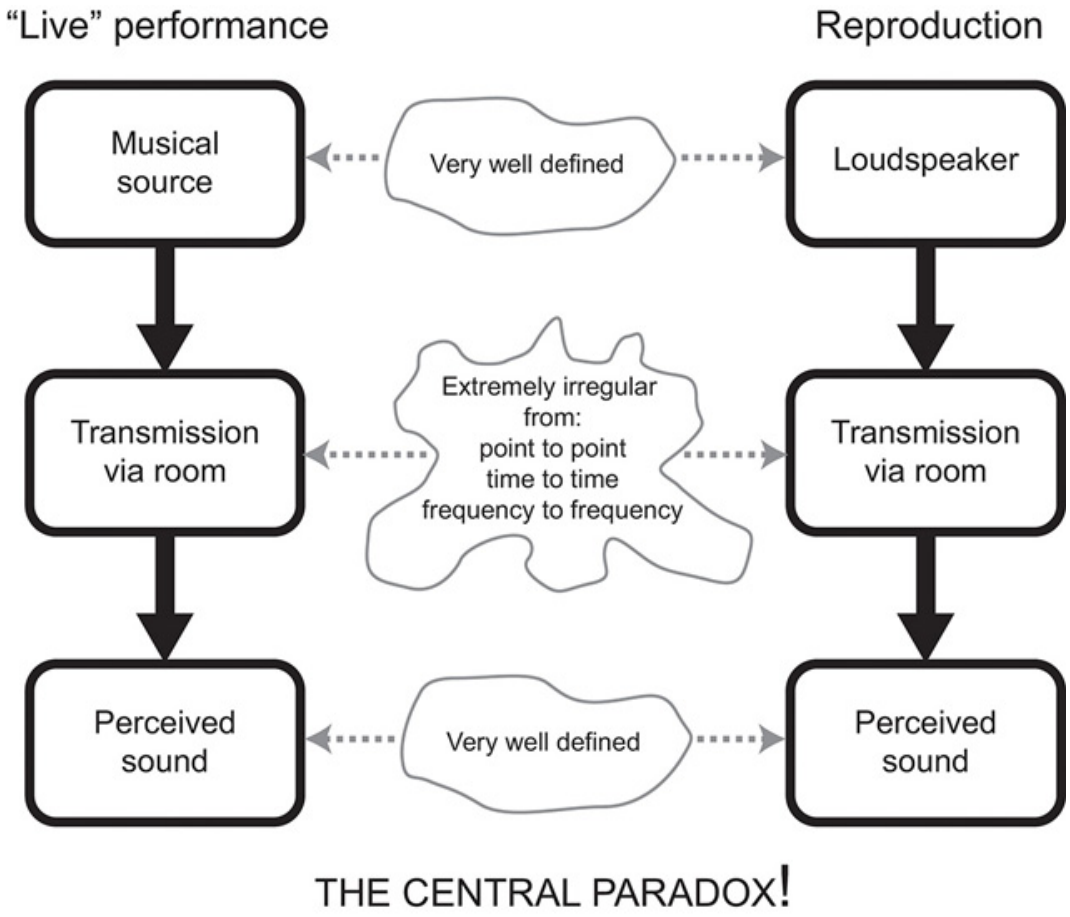


FIGURE 1.6 *The "Central Paradox" of live sound in rooms inspired by Benade (1984), expanded by the author to include reproduced sound.* ⏎

In this book the focus is on the right side of Figure 1.6—sound reproduction—where loudspeakers are expected to reproduce captured sounds from voices and musical instruments. It seems logical that there are parallel experiences to those in live performances, where listeners are able to "listen through" small rooms to discern timbral subtleties of reproduced sounds. The fundamental difference between the two parallel paths is that in the "live" path on the left we tend not to question the sound quality of the performance—the performance itself, perhaps. It is the ultimate "sighted" listening test. All biases are in place. In contrast, on the right "reproduction" path questioning the sound quality is expected, and if we don't like it, we want it fixed. See OR Chapter 1 for insight into how musical instruments and live performances are judged. It is difficult to eliminate biases.

An important fact is that the performance of loudspeakers can be measured, and if we know what we are doing, we may be able to determine how accurately they can reproduce the sounds of voices and instruments. This is ultimately judged by listening to them and forming opinions in small rooms.

The apparently simple concept illustrated in Figure 1.6 leads to speculation about what it is that listeners are influenced by. It is reasonable to believe that *up to a point* we "listen through" rooms to discern subtle qualities in the sounds of voices and musical instruments. But there are limits. It is the purpose of research to identify the limits, how to measure them, and how to attenuate the components in the sound field that exceed the limits. Eliminating the room is not the answer. Listening in a reflection-free space, like an anechoic chamber, is not pleasant. It is not pleasant even to converse in. We need some room sound, but how much, and what kind?

1.6.1 Should We Measure the Output of Loudspeakers or the Input to the Ears?

The traditional approach to measuring loudspeakers has been to measure as much of the sound output as possible, from as many directions as possible, and find ways to present this three-dimensional data in ways that are informative. This is rational, and it has led to some very creative and

colorful data presentations. But, really, what we need to understand is what portions of that total sound arrive at the listening position in sufficient strength to influence what is heard. The latter data are a subset of the former, probably post-processed.

In professional audio, where complex arrays are used to address large audiences embracing many angles and distances, sound output as a function of direction is critical information. Transfer function (amplitude and phase/impulse) data with high angular resolution is necessary to allow computer models to predict the sound pattern radiated by combinations of loudspeaker modules in the enormous arrays we see at outdoor concerts and large indoor venues. It is advanced computer modeling and engineering, and the required data files are enormous.

Listening to smallish loudspeakers at short distances in small rooms presents a very different situation, and the measurement requirements are different. The floor and ceiling are likely to be horizontal surfaces, and the walls vertical, so the planar orientations of the strongest reflections are known for most listening situations. Therefore, all sounds radiated in all directions from the loudspeaker are not equally important to the listener.

The explanations will come later, but there is merit in estimating the contributions of direct and early reflected sounds and sound power separately, employing statistical geometric data on typical domestic listening circumstances to identify the appropriate off-axis measurements. We will be looking at loudspeaker performance from the listener's perspective. Listeners are the final judges of what constitutes good sound quality, so it is reasonable to try to understand what sounds arrive at their ears—what are they hearing? This is a departure from traditional loudspeaker measurement methods.

1.7 CIRCLES OF CONFUSION WITHIN THE RECORDING INDUSTRY

Musicians and engineers in recording control rooms and mastering rooms listen to loudspeakers while creating the art, so improvements in

understanding loudspeakers and rooms should have a beneficial effect on the art itself. The "circle of confusion" shown in Figure 1.7 illustrates a perspective.

Figure 1.7 describes the conundrum created when subjective opinions are all one has. I called it a circle of confusion, with apologies to the optical industry where the term originated. It explains that loudspeakers produce sounds allowing us to form opinions of microphones that are used in making the recordings we use to evaluate the microphones by listening through loudspeakers. The "psychoacoustic" clock detonator and dynamite "measurements" imply that there will be a time when we understand enough psychoacoustics about the meaning of measurements that they will explode the circle of confusion.

Later, the circle was modified to apply to the recording industry more directly, as shown in Figure 1.8.

Figure 1.8 illustrates the circle of confusion that exists within the audio industry. It is simplified to focus on the loudspeaker, but the room is also a factor of course—especially in the low-frequency resonance region below about 400 Hz. The basic circle starts with monitor loudspeakers that are evaluated using recordings that are made using microphones, equalization and other effects that are evaluated using monitor loudspeakers—the circle is complete. The resulting recordings are then used as test signals to subjectively evaluate audio equipment, as well as being enjoyed by consumers in countless listening situations. The characteristics of the monitor loudspeakers must influence the sound of the recordings, which means that unless consumers have similar sounding loudspeakers, they won't hear the recorded art as it was created.



FIGURE 1.7 *The author's first attempt at illustrating the awkward situation confronting the recording industry—a circle of confusion.*

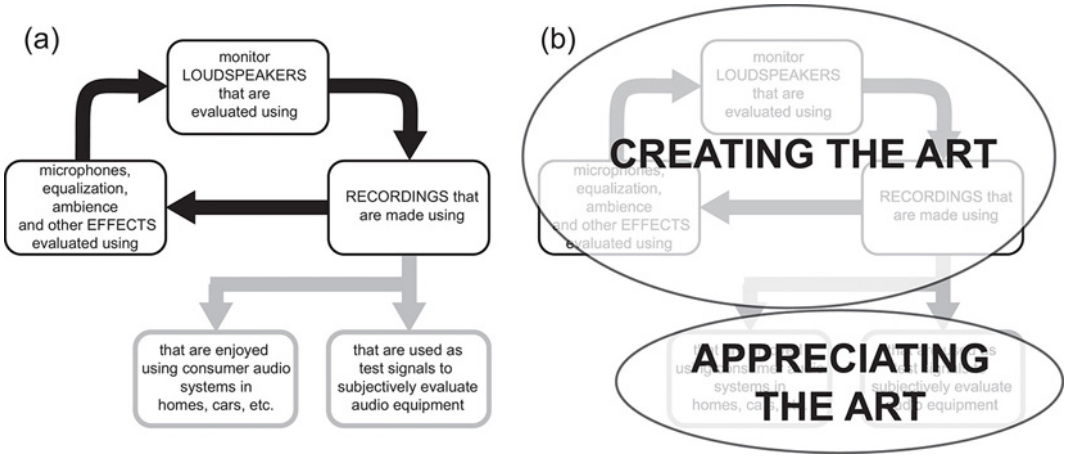


FIGURE 1.8 *(a) The circle of confusion as it relates to the audio industry showing the important role of loudspeakers in determining what recordings sound like. (b) Unless there are significant common technical factors in the domains where the art*

is created and where it is appreciated, the art is not preserved; it does not perfectly "translate" in professional audio terms. ⏎

The professional domain within which the art is created is both technically and subjectively complex. We will return to this topic throughout the book, as it is central to the audio industry as a whole. The art created by talented composers and musicians, crafted by skilled recording engineers, should have a reasonable chance of being heard, as created, by the consuming public. It should translate. Translation will be discussed in some detail in Section 12.5, and there are many variables.

Over the decades, consumers have lived with, and professionals have worked with, loudspeakers having serious sound-quality flaws, yet recordings have been made, reproduced and enjoyed. Good music is very durable. Humans are very adaptable. But time has passed, and now loudspeakers exist that are fundamentally neutral, transparent reproducers of sound, and they exist in versions addressing professional and consumer markets. It is sensible that they be used in both domains. Many are, but not universally, which adds to the problem of how recordings "translate" to listeners in the outside world.

Some engineers say they must "like" the monitor loudspeaker. It is fundamentally important to like what is heard, but is it the loudspeaker or the program? Obviously, it is the combination. The program is the art that is being created for distribution to the world. Loudspeakers are technical devices that transduce electrical signals into sound. We know how to measure them. It is possible to identify loudspeakers that are the most transparent transducers of the electrical signals—as subjectively judged in double-blind evaluations. This was not always possible, and not all manufacturers are yet ready to reveal this kind of data. Independent sources exist, fortunately.

Incredibly, there are people who argue that timbrally neutral loudspeakers are boring, robbing music of some undefined inspirational qualities. This, of course, is nonsense. It is the music, the art, the recordings that should inspire, excite, please and motivate listeners. Loudspeakers are messengers

of the art, and they need to be honest messengers for both professionals and consumers. If anyone wants to add extra inspiration, excitement or more apparent reality there are signal processors and equalizers to modify the spectrum and space. For consumers, starting with neutral loudspeakers allows them to return to accuracy at any time. Recording engineers need neutral references to hear what their electronic manipulations sound like to increasing numbers of consumers.

Headphones are very different devices, and they totally avoid the room-acoustics complications. Headphone listening has a fundamental problem in that the soundstage and imaging are designed for and best appreciated through a pair of loudspeakers. Through headphones these aspects are greatly altered. In compensation, there is a clarity and immediacy to the sound that is lacking in conventional loudspeaker stereo. Headphones are a rapidly developing technology, and fortunately we are now able to identify the basics of timbral neutrality in the right kinds of measurements. The topic will be discussed in a later chapter.

This book describes the scientific research conducted by Floyd Toole and collaborating co-authors Sean Olive and Todd Welti over several decades. It also includes research done by many others around the world. The accumulated science provides foundations for guidelines for designing, selecting, and using loudspeakers and headphones likely to deliver minimally altered reproductions of recorded music and movies.

Although there are still things to be learned, much is now known, and the result is better sound in more places. Read on.

REFERENCES

ANSI/CTA-2034-A R-2020 (2015). “Standard Method of Measurement for In-Home Loudspeakers”, *Consumer Technology Association, Technology and Standards Dept*. www.CTA.tech. ⏎

Benade, A. H. (1984). “Wind Instruments in the Concert Hall” Text of an oral presentation at Parc de la Villette, Paris; part of a series of lectures

entitled “Acoustique, Musique, Espaces”, May 15 (personal communication). ⏎

Blauert, J. (1996). *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, MA. ⏎

Bregman, A. S. (1999). *Auditory Scene Analysis, the Perceptual Organization of Sound*, MIT Press. ⏎

Cherry, Colin (1980). *On Human Communication*, 3rd edition, The MIT Press (originally published in 1957) ⏎

Helson, H. (1964). *Adaptation-Level Theory*, Harper and Row, London. ⏎

Pike, C., Mason, R. and Brookes, T. (2014). “The Effect of Auditory Memory on the Perception of Timbre”, *136th Convention, Audio Eng. Soc*., Paper 9028. ⏎

Sayers, B. McA. and Toole, F. E. (1964). “Acoustical Image Lateralization Judgments with Binaural Transients”, *J. Acoust. Soc. Am*., vol. 36, pp. 1199–1205. ⏎

Schuck, P., Olive, S., Toole, F., Sally, S., Bonneville, M., Momtahan, K. and Verrault, E. (1993). “Perception of Perceived Sound in Rooms: Some Results of the Athena Project”, *12th International Conference, Audio Eng. Soc*., Paper 12–006. ⏎

Toole, F. E. (1982). “Listening Tests—Turning Opinion into Fact”, *J. Audio Eng. Soc*., vol. 30, pp. 431–445. ⏎

Toole, F. E. (1998). “Audio—Science in the Service of Art”. www.harman.com/documents/AudioScience_0.pdf ⏎

Toole, F. E. (2015). “The Measurement and Calibration of Sound Reproducing Systems”, *J. Audio Eng. Soc*., vol. 63, pp. 512–541. This is an open-access paper available to non-members at www.aes.org www.aes.org/e-lib/browse.cfm?elib=17839 ⏎

Toole, F. E. and Olive, S. E. (1994). “Hearing Is Believing vs. Believing Is Hearing: Blind vs. Sighted Listening Tests and Other Interesting Things”, *97th Convention, Audio Eng. Soc*., Preprint 3894. ⏎

Toole, F. E. and Sayers, B. McA. (1965a). “Laterization Judgments and the Nature of Binaural Acoustic Images”, *J. Acoust. Soc. Am*., vol. 37, pp. 319–324. ⏎

Toole, F. E. and Sayers, B. McA. (1965b). “Inferences of Neural Activity Associated with Binaural Acoustic Images”, *J. Acoust. Soc. Am*., vol. 37, pp. 769–779. ⏎

Zielinski, S., Rumsey, F. and Bech, S. (2008). “On Some Biases Encountered in Modern Audio Quality Listening Tests—A Review”, *J. Audio Eng. Soc*., vol. 56, no. 6, pp. 427–451 ⏎

CHAPTE
R 2

A New Standard for Loudspeaker Measurements

Floyd Toole

DOI: 10.4324/9781003477495-3

After it was decided that the research would focus on loudspeaker sound quality, an existing National Research Council of Canada laboratory was converted to a listening room. It was shortened to improve the room mode distribution. Because it had concrete walls and floors it was far too reflective, so custom-designed membrane absorbers attenuated strong low-frequency room modes, and resistive absorbers brought frequency-dependent reverberation times to values typical of domestic dwellings (Toole, 1982). Custom signal-switching hardware made it convenient to set up and conduct multiple-loudspeaker (three or four at a time), equal-loudness double-blind comparisons. Many such tests were performed, providing a basis for subjective/objective comparisons and correlations. On each loudspeaker one set of technical measurements is sufficient, but

subjective tests must be repeated and randomized, and must involve several listeners to be statistically useful data. Subjective data takes facilities, time and labor to acquire, which is why the demanding listening tests are often avoided or overly simplified.

I attended several International Electrotechnical Commission (IEC) standards meetings on loudspeaker listening tests. I had hoped to learn from others on the committee but found that my own work was the main topic of conversation. It was sufficiently respected that the NRCC room was chosen as the prototype for a new international recommended practice: IEC 268–13 "Listening Tests on Loudspeakers" 1985. I wrote large portions of that standard. The original document envisaged the construction of "standard" listening rooms that would allow for comparisons of experimental results obtained within them. Only a few such rooms were built; expense was a factor. Later editions of the document relaxed the requirements to the point where many rooms would qualify as meeting the requirements, eroding the key point of the recommendation but pleasing those wishing to be able to justify use of their existing rooms (IEC 60268-13 1998).

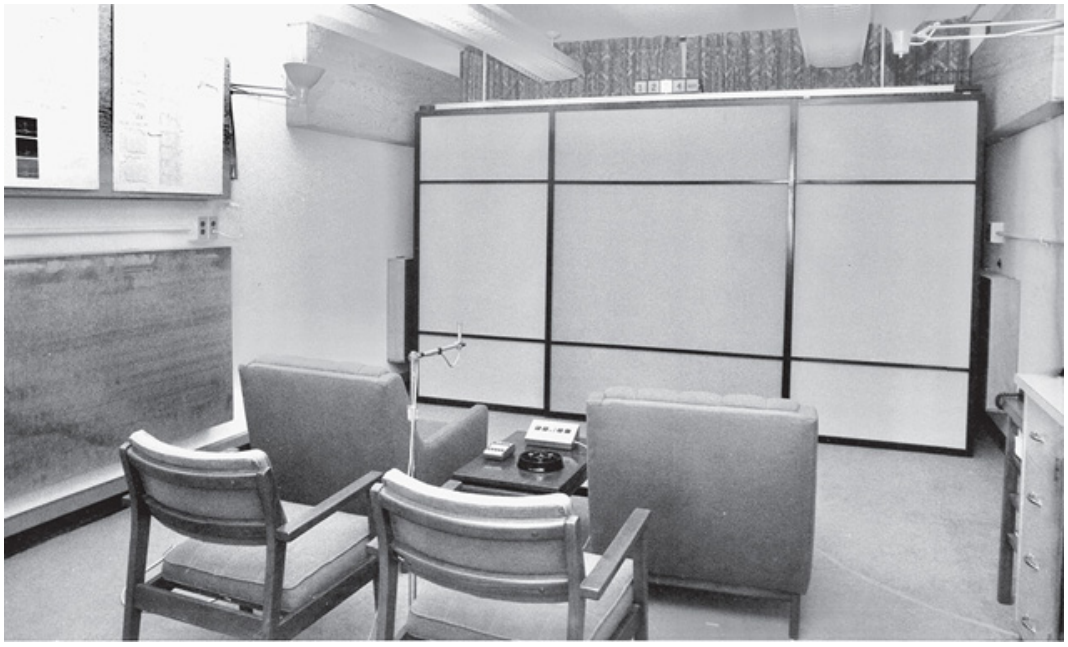


FIGURE 2.1 *The NRCC listening room in an early configuration for monophonic comparisons of multiple*

loudspeakers. The screen was acoustically very transparent. The microphone between chairs was used to adjust for equal loudness and overall sound level. ⏎

In retrospect, we now know that room resonances at low frequencies account for about 30% of overall evaluations of sound quality, so adhering to the new standard is no assurance of obtaining "standardized" subjective data. Room modes will be different for different IEC rooms, so some form of control or normalization is required below about 300 to 500 Hz to evaluate what is happening at higher frequencies. This topic will be addressed in Chapter 14.

Figure 2.1 shows a photograph of the original NRCC listening room set up for monophonic, multiple-loudspeaker evaluations. Four chairs were arranged around a small table on which is a switchbox for selecting loudspeakers, and a microphone at ear level for setting equal loudness and listening level. The loudspeakers were hidden behind an acoustically transparent screen.

Listeners switched among the loudspeakers at will. Above the screen is a display showing which numerical loudspeaker is active. The number identifiers were randomly changed during the multiple listening sessions, as were the loudspeaker and listener locations to randomize the effects of low-frequency room standing waves (resonances). The loudspeaker locations and music changes were predetermined using random number guided sequences to avoid experimenter manipulation of results—this is the "double" part of double blind. For stereo evaluations two chairs were arranged on the center line, with the rear chair higher than the front chair.

Having on-site capability to perform reliable, repeatable subjective and objective evaluations as needed greatly facilitated the research. From the outset, local enthusiasts and audio stores eagerly provided loudspeakers for testing and ears for listening. It was interesting that some of the audio salesmen found that participating in the unbiased product evaluations interfered with their abilities to sell certain profitable products, and they voluntarily retired from listening panels. Business is business.

A major contribution to the research came from Canadian audio magazines *Electron, AudioScene Canada*, and *Sound Canada*, which published articles by me interpreting science to laypersons and describing some of the research findings. Ernie Welling, Ian Masters and Alan Lofft were editors of magazines that paid the NRCC to have products measured. They and other journalists traveled to Ottawa to participate in double-blind listening tests. They saw the measurements only after the subjective evaluations were done. This provided a continuing flow of products, and further publicity. The resulting magazine reviews were exemplary in their candor, and anechoic measurement data were shown—an industry first? (See OR Chapter 2.) But manufacturers (advertisers) do not appreciate negative reviews, so there was sacrifice involved in being honest.

Some American audio publications made enquiries, but border complications and cost were impediments. Surprisingly, Harry Pearson, editor of *The Absolute Sound*, a US publication known for its subjectivist approach to audio, showed great interest. Through his connections in Canada, he had been following our work and asked for exclusive access to the review services. This could not be granted because the NRCC is a Canadian government facility, but it is intriguing to wonder what might have become of the publication if the arrangement had been realized. Harry clearly appreciated the value of science, but publishing a profitable magazine is a different thing, and the editorial direction did not change.

2.1 EXAMINING THE MUSIC SOURCE OF THE DAY: LPS

In parallel with the loudspeaker research, the performances of LPs, turntables and phono cartridges were investigated, as these defined the prevailing source of test music. The goal was to maximize the performance of the medium, but what was learned was that the technology was so fundamentally flawed that LPs were incapable of reproducing the sounds on master tapes, the original art. Rumble, wow, flutter, and surface noises of assorted kinds were familiar problems, but acoustical pickup from unsupported disc surfaces, and vibration feedthrough from inadequate

turntable suspensions were less well known. Some styles of tone arms were prone to mechanical resonances (see OR Chapter 2).

Tracing distortions were significant. LP masters are cut with a flat fronted chisel and then played with a conical or elliptical stylus, meaning that the path followed by the stylus is different from that of the cutter. The cutter also follows a radial straight-line path, while pivoted tone arms scan an arc, causing misalignment over much of the recorded surface. Signal manipulations during the disc mastering process attempt to compensate for some of these, including amplitude compression and spectral manipulations to avoid stylus mistracking (momentarily losing contact with the groove walls) in the inner grooves where classical music often reaches a climax. The bass in both channels is usually combined to mono, allowing horizontal motion only, to prevent the stylus from being thrown out of the groove. Disc mastering is a complex art, trying to preserve as much of the master tape performance as possible.

RCA "Dynagroove" LPs were pre-distorted in an attempt to cancel tracing distortion, but that distortion is different for styli of different shapes, so results varied. If that were all that was done, it might have had some positive effect, but the process also added bass at low levels, treble at higher levels and made dynamic range modifications intended to "improve the listening experience". It met with mixed reviews and was generally not attractive to purists.

To better understand the situation, I collaborated with Alan Lofft, editor of *Sound Canada* magazine, and Dr. Wieslaw Woszczyk at McGill University's sound recording program to create a test LP, with technical signals and musical excerpts from master tapes. The LP mastering process is critical, so they used the mastering lab that did the highly regarded CBS test records. The comparisons between LP reproduction and the master tape were not flattering, and some non-linear distortions measured in whole percentages. It is remarkable how effective perceptual masking is in tempering the distracting effects of these distortions and surface noise. In this research we were not listening for pleasure; the music was a test signal and needed to be as close to the original master tape as possible.

A prominent British classical record label courageously provided a one-from-master tape of symphonic music and the corresponding LP for evaluation, and again it was clear that they were not the same. From that point on, one-from-master tapes were used whenever possible, until digital sources were available. Today, digital recordings of analog masters are indistinguishable from the originals, and digital master recordings are the references. Still LPs have followers, some because of nostalgic attachments, cover art and the attractions of a playback system that needs personal involvement. LPs have a distinctive “sound”, and it is not the sound of the master tape, the original art.

A sad fact: some people responsible for master tape archives obviously felt that LPs were the final stage of audio evolution and sometimes, to save expensive tape, allowed original master tapes to be reused, preserving only the pre-processed LP cutting tapes. Obviously, any CDs made from these tapes were corrupted at the source, explaining some disappointing experiences with the new medium when LP and CD versions of the same recording were compared.

2.2 THE RESEARCH AIDS MANUFACTURERS AND A BROADCASTER

As a result of the magazine publicity and word of mouth, designers from emerging Canadian loudspeaker brands (PSB, Paradigm, Axiom, Energy, Mirage, Definitive Technology, and others long gone) rented the NRCC loudspeaker facility to develop prototype loudspeakers and to subjectively evaluate them using the double-blind method. All this subjective and objective information added to the database for the ongoing research. Even though they were competing against each other in the marketplace, these young companies had evidence that they were ahead of much of the international competition in terms of sound quality. There was a definite camaraderie among them. Of course, sound quality alone does not guarantee sales, but it cannot hurt. The scientific research and tours of the

laboratory facility gave these brands a marketing story and instant credibility in the audio world.

Word of the research had spread to the Canadian Broadcasting Corporation, a Crown corporation operating nationwide radio and television networks in both English and French. The engineering department approached me for assistance in selecting monitor loudspeakers for use throughout the organization. They wanted large, medium and small monitor loudspeakers that were good sounding and similar in sound. They would provide the loudspeakers and several experienced listeners from their recording engineering and production staff. It was an opportunity I could not refuse.

The submitted samples from well-known brands and multiple countries were measured, and several were eliminated immediately based on obviously unacceptable performance, including mismatched left and right samples—not a good sign. The remainder were incorporated into carefully randomized double-blind, equal-loudness, subjective evaluations that went on for several days. Along with the professional monitor samples, I added a few Canadian consumer loudspeakers that had been highly rated in my regular evaluations. A few of my regular listeners participated.

The results had some surprises. First, the preferences of the professional listeners did not differ from those of my regular listeners, which was encouraging. But the most serious and significant observation was that several of the professional listeners were noticeably inconsistent in their ratings of the same loudspeakers in the randomized repeated exposures. Everyone exhibits variations, but these listeners stood out because of the magnitude of the variations. They could like a product in the morning and dislike it in the afternoon, writing articulate explanations at each sitting. Looking for a cause, it was quickly revealed because we had decided to perform audiometric tests on all listeners beforehand. In short, even modest hearing loss resulted in measurably increased variability in judgments of sound quality. We all came to realize that hearing loss is an occupational hazard in professional audio, which is most unfortunate in that these people

decide what recordings sound like. This will be discussed in Section 10.1.2 and Chapter 16.

The other surprise was how poorly many of the professional monitors were rated, compared to the consumer products that had been included. Remember, this was a rating of sound quality, not loudness capability and power handling—separate issues. Two engineers commented that they had never heard better sound in their lives than some they heard during the evaluations. But other engineers were upset because they had disliked their favorite monitors and insisted on a retest using their own master tapes. This was done and the results were the same. When they saw the measurements, and had them explained, there was acceptance, but it was yet another example of habituation. Humans are good at adaptation—normalizing unusual things.

The exercise was successful, monitors were selected for each category, and it was gratifying to see some Canadian consumer loudspeakers being approved for moderate loudness applications.

2.3 INFORMATIVE TECHNICAL MEASUREMENTS: MAJOR PROGRESS

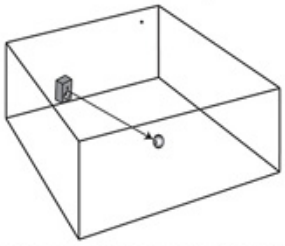
Following the decision to commit to the loudspeaker investigations, the NRCC anechoic chamber was reconfigured for convenient loudspeaker measurements, starting with a rotating platform allowing full 360° measurements at 2 m distance. In the beginning the data were acquired using the then state-of-the-art Brüel & Kjær analog measurement system, generating accurate ink plots on graph paper. Post-processing of data was very limited, and manual, but despite the difficulties, one could see that off-axis measurements were valuable data. In 1983 the analog system was replaced by a custom hybrid analog/digital system controlled by a DEC PDP-11/03 computer that permitted storage and post-processing of data acquired with 1/20-octave resolution. Low-frequency calibration of the chamber was then possible. It was a breakthrough.

The floor, ceiling and walls in most listening rooms are horizontal and vertical, and these provide the strongest first-reflected sounds. All other reflections travel farther (more inverse-square loss and air attenuation) and encounter multiple reflecting surfaces, losing energy at each contact. Consequently, measurements were made on horizontal and vertical orbits at 15° intervals in the front hemisphere and 30° intervals in the rear hemisphere; 34 measurements in all. Comparisons, sums and differences of curves, spatial averages of multiple curves and total sound power output estimates were then possible. It was a major step forward in extracting information from the raw data.

Figure 2.2 explains why normal room geometry determined why the anechoic chamber apparatus for measuring loudspeakers focused on sounds radiated on horizontal and vertical orbits.

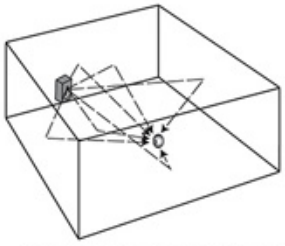
Figure 2.3 shows the physical setup in the NRCC anechoic chamber. As shown, it measures the horizontal axis data. The loudspeaker is then placed on its side, with the reference point aligned over the axis of turntable rotation and elevated to align with the microphone. Small loudspeakers are easy to arrange for these measurements, but tall, floor-standing loudspeakers with woofers at the bottom present the structural challenge of supporting the mass, most of which is located at a distance from the rotational axis. The structure must be strong, with facilities for clamping tall loudspeakers.

The first-arriving direct-sound dominates sound localization and significantly influences perceived sound quality



The direct-sound is well described by the on-axis anechoic frequency response when the loudspeaker is aimed at the prime listener.

Early-reflected sounds are the second loudest sounds the listener hears. They influence most things we hear, and dominate in-room measurements.



Because rooms have vertical walls and horizontal floor and ceiling, the strongest sounds are measured on horizontal and vertical orbits.

FIGURE 2.2 *The propagation paths of direct and early-reflected sounds show that two of the most important measures of loudspeaker performance do not need full 360° spherical anechoic data.*

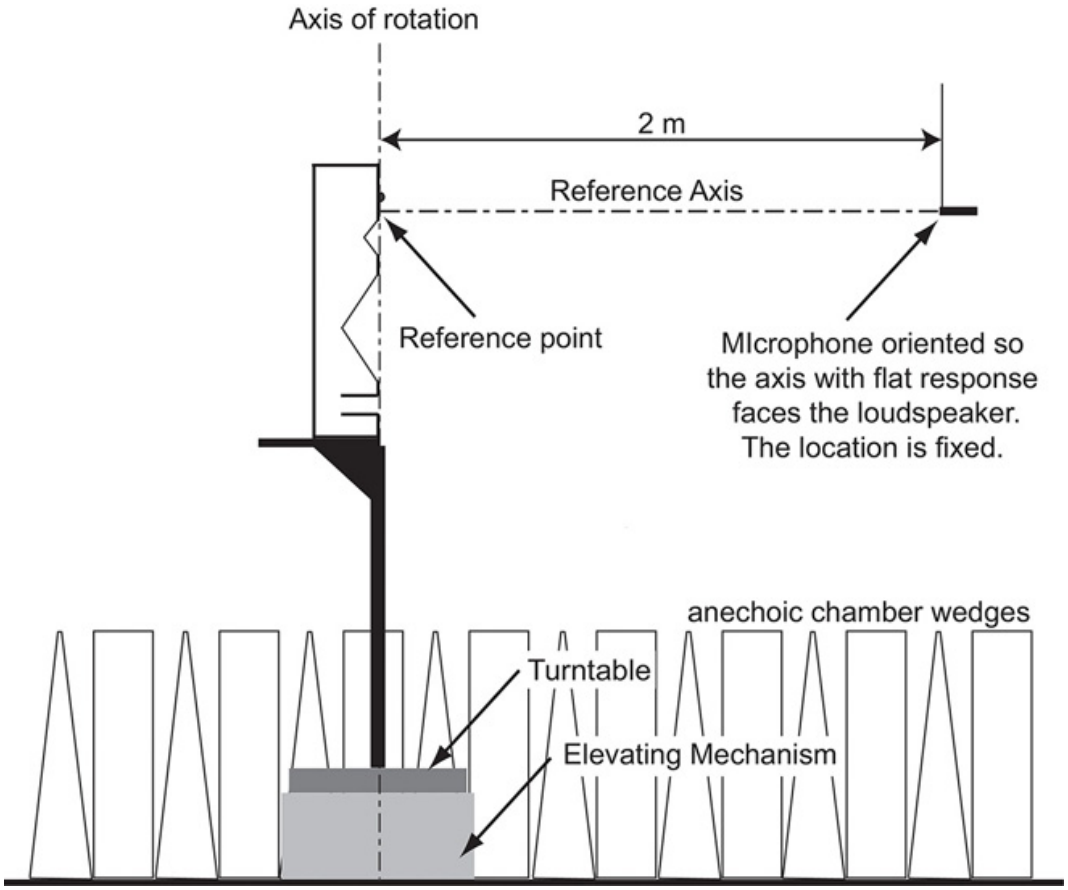


FIGURE 2.3 *The measurement configuration in the NRCC anechoic chamber, showing the test loudspeaker on a platform that rotates it around a reference point on the front baffle. The height of the platform is adjusted to align the reference point with the height of the measurement microphone at 2 m distance.*

The low-frequency calibration of the chamber presents a complication. Anechoic chambers cease to be perfectly echo free at very low frequencies, and errors arise. The NRCC chamber had 1 m long wedges, and a space behind them, so it was usefully anechoic down to about 80 Hz. To

compensate for errors at lower frequencies, reference measurements were made using a closed-box woofer on a 10 m tower and then the digital post-processing made corrections so that the same woofer located at the axis of rotation in the chamber yielded the same result. Henceforth, any loudspeaker with a similar bass configuration would have free-field corrected bass measurements.

With tall, floor-standing loudspeakers placed horizontally the woofers scan an arc a distance from the calibration point/axis of rotation. This means that they do not interact with the residual standing waves in the chamber in the same way that the calibration woofer did. On any measurement axis there may be low-frequency calibration errors due to this and to the placement of reflex ports. A pragmatic solution was to use the sound power calculation, which involves all 34 measurements, as a reasonable estimate. With the profound effects of small room resonances and adjacent boundaries these measurement errors are unlikely to be factors in real-world audible performance. Relocating the woofers of tall loudspeakers to be closer to the axis of rotation can yield more precise data at low frequencies, which then can be spliced to the mid-/high-frequency data. Ideally, every distinctive configuration of loudspeaker would need to have a reference curve measured on a tower, or ground-plane equivalent, and I know of one manufacturer that takes the trouble to do this.

A multiple-microphone array was installed to allow rapid evaluation of on-axis and listening window frequency response curves. The listening window was, I believe, a new concept at the time, providing an average of measurements at ±15° vertically, and ±15° and ±30° horizontally, thereby estimating the average direct sound likely to be heard by members of an audience. The spatial averaging also attenuated visual evidence of acoustical interference, better revealing the audible character of direct sound.

2.3.1 Early Results Were Very Revealing

Being able to see families of full-bandwidth, high-resolution frequency response curves in a single presentation was a huge advantage. It allowed

one to visually recognize patterns in the collection providing insights into how physical performance attributes related to subjective perceptions. Spatial averaging of multiple curves allowed for the visual separation of relatively innocent acoustical interference effects from those associated with easily audible resonances. A single curve is insufficient information.

The anechoic measurements from that system were post-processed to display the following indicators of linear behavior:

1. On-axis amplitude vs. frequency (frequency response)
2. Individual curves at each of the 33 off-axis measurement locations
3. Mean amplitude response in the front hemisphere
4. Mean amplitude response in the total sphere
5. Mean amplitude response in a ±15° "listening window" (later expanded to ±30° horizontal)
6. Mean amplitude response in the ±30°–45° annulus (including horizontal and vertical data)
7. Mean amplitude response in the ±60°–75° annulus (including horizontal and vertical data)
8. Total sound power calculated from individual measurements weighted according to their contributions to the total power radiated through a spherical surface
9. Directivity index
10. Phase response

More details can be found in sections 5.2 and 5.3 in Toole (1986, part 2). The remainder of that paper shows many examples of the measurements displayed according to the subjective ratings of sound quality by listeners in double-blind tests.

There is no doubt from these displays that there was an intimate relationship between the smoothness and flatness of these curves and the perceived sound quality (then called "fidelity") ratings of the loudspeakers. *Fidelity* was rated on a scale of 10, where 10 represents a reproduction that is perfectly faithful to the ideal, no improvement is possible, and 0 represents the worst reproduction imaginable. A *Pleasantness* scale was also provided, thinking that there might be aspects of sound that were appealing, but not accurate. The scale was eventually dropped because the two ratings simply tracked one another. The tests involved double-blind randomized comparisons of four loudspeakers at a time, with several repeated presentations of the same loudspeaker in different groupings.

Listeners were professional colleagues, friends interested in audio, and a few audio journalists. Some were amateur musicians. None was a "trained" listener, but all who participated on a regular basis became very adept at the task. Switching among the loudspeakers was instantaneous. Many tests involved single listeners—the best method. Others involved two or three at a time who shared control. This was not ideal, because the switching sequence quickly revealed which loudspeakers the person in control was favoring and which were being ignored. An independent operator following a randomized sequence, and switching at suitable intervals was preferred. Several musical selections were used—classical, jazz and popular. There was no time limit, but it became clear that, with experience, decisions were made quickly. A selection of programs that listeners found useful also emerged. Figure 2.4 shows an example of the results.

Some trends are worth noting in Figure 2.4:

- There is an underlying "flat" trend in these clusters of frequency-response curves. The variations, even the larger ones, seem to be fluctuations around a horizontal line for the on-axis groups, and around gently sloping lines for the off-axis groups. Imperfections can be seen in the crossover regions of several of the curves. Some are related to mismatched transducer directivity at the crossover frequency, others to non-optimum network design.

- The low-cutoff frequency progressively decreases as the fidelity rating increases. The listeners responded favorably to bass that extended to lower frequencies, not boosted bass.
- The best-sounding loudspeakers all had smooth, gently undulating phase responses. Lower ratings were given to loudspeakers with resonances that showed up as discontinuities in the curves; high-Q resonances as sharp discontinuities. The most useful evidence of resonances is repeated patterns in the family of frequency response curves. The lack of any persuasive visual correlation between phase response and fidelity rating caused it to be relegated to the category of "optional" measurements. Of course, for system design, phase behavior is essential to the creation of properly functioning crossovers. The contributions from the low-pass and high-pass portions must acoustically sum smoothly.

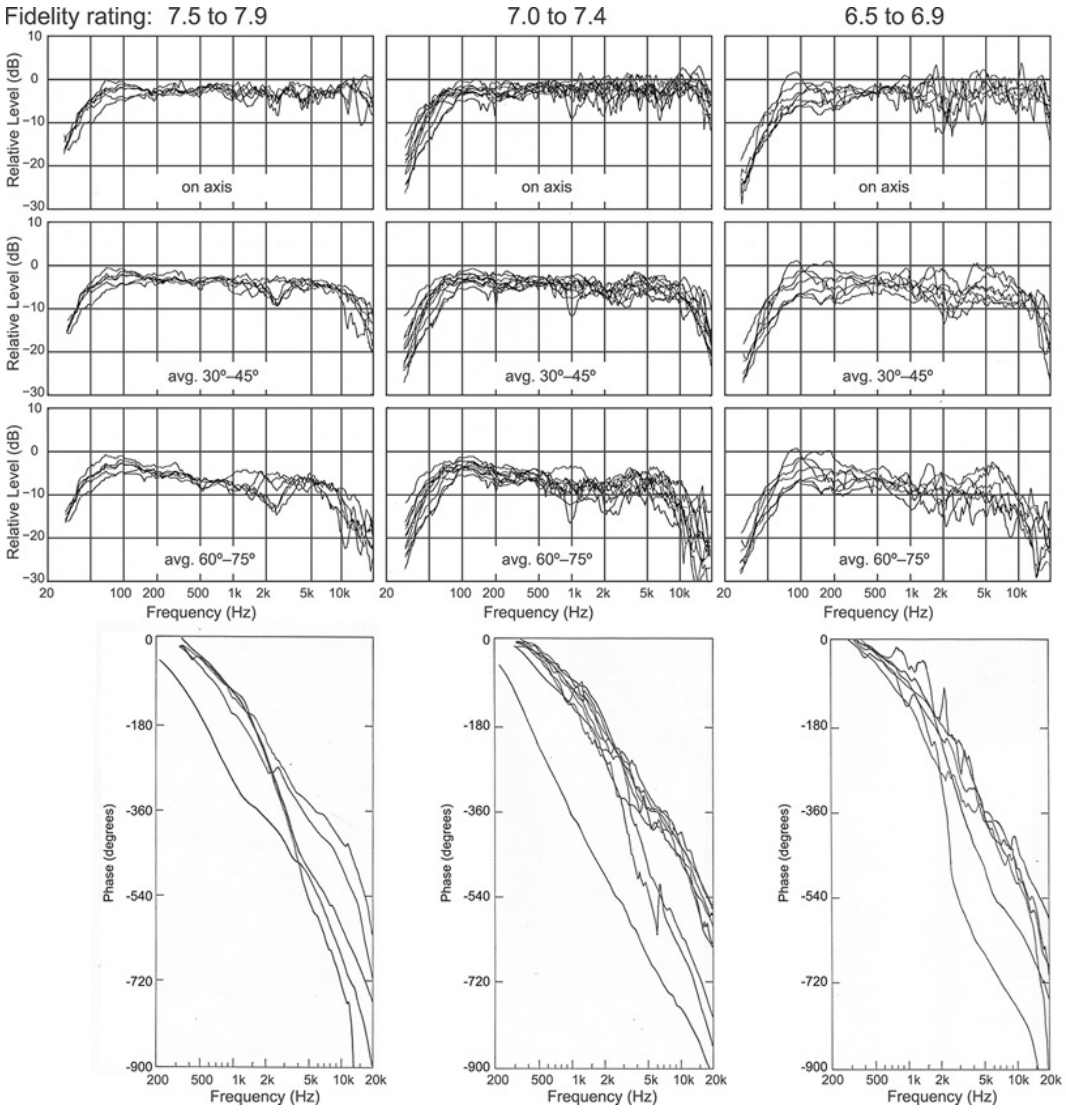


FIGURE 2.4 *A sample of anechoic measurement results from Toole, 1986, part 2, figures 7 and 13, showing loudspeakers grouped according to subjective "fidelity" ratings in three categories. There were 6 loudspeakers awarded ratings between 7.5 and 7.9, 11 loudspeakers in the range 7.0 to 7.4, and 7 loudspeakers in the range 6.5 to 6.9. The original data include a fourth, lower, category that adds nothing to the discussion. The measurements are unsmoothed, 200-point, log-spaced stepped-tone anechoic measurements (1/20 octave). To eliminate the effects of loudspeaker sensitivity the vertical positions of the frequency-response curves were normalized to the mean sound*

level in the 300–3000 Hz band. This same frequency band was used with pink noise to normalize listening levels during subjective evaluations. The phase responses may include a small residual time-delay error that could affect the slopes of the curves but not the contours and fine structure. ⏎

Clearly the commonly used ±3 dB "acceptable" tolerance does not do justice to the highest-rated (7.5 to 7.9) loudspeakers. It is evident that smooth and flat was the design objective for all these loudspeakers, and listeners could hear it. The variations seen among these six loudspeakers from different designers, manufacturers and countries of origin are smaller than the production tolerances allowed by some manufacturers for a single model. Thirty-nine years later it would not be embarrassing to compare these loudspeakers to comparable products in today's marketplace.

The apparent preference for extended bass motivated Figure 2.5, where the low-cutoff frequencies for all of the loudspeakers were determined at two attenuation levels, −5 dB and −10 dB, relative to the reference 300–3000 Hz band. The normal −3 dB "half-power" level was also tried, but it yielded no relationship, which was anticipated because of the substantial effects of solid angle gains and bass resonances in small rooms. Therefore, it was not a total surprise to find that the relationship between "fidelity" rating and low-frequency cutoff reached a maximum correlation for cutoff frequencies determined at the −10 dB level.

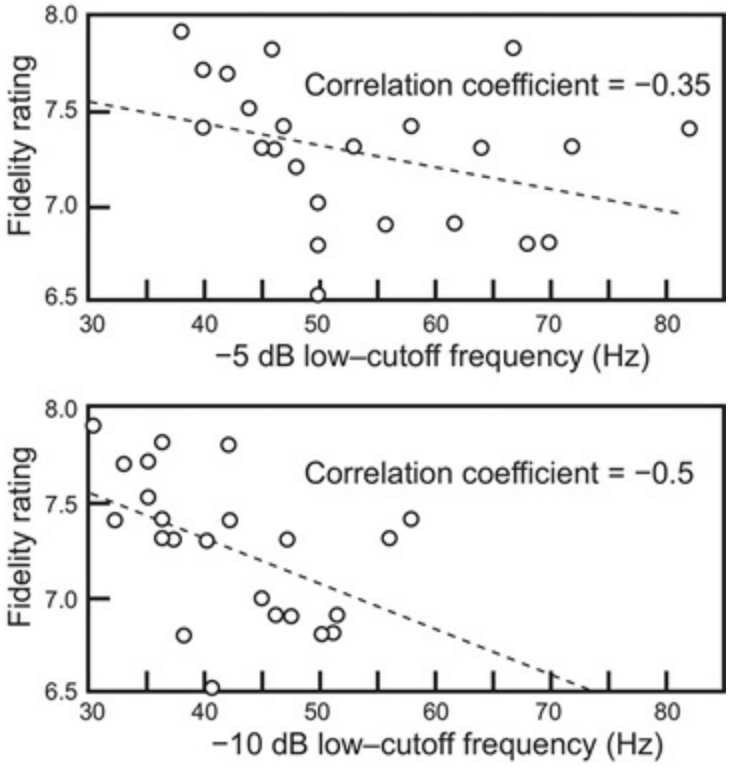


FIGURE 2.5 *Distribution diagrams of fidelity ratings vs. low-cutoff frequency for −5 dB and −10 dB reference levels compared to the average sound level over the 300–3000 Hz band. Regression lines and correlation coefficients indicate clear trends, especially for the −10 dB data. From Toole (1986, part 2, figure 10).* ⏎

Bear in mind that this is a correlation achieved with all other factors varying, indicating that bass extension is a *very* important factor in overall sound quality evaluations. This was confirmed many years later in work by Olive (2004a, 2004b), discussed later, which revealed that bass performance —smoothness and extension—accounted for about 30% of the factor weighting in double-blind loudspeaker evaluations of sound quality.

These results were extremely gratifying, indicating that given an unbiased listening situation, listeners agreed on which loudspeakers sounded “good”. The preferences exhibited a simple visual correlation with the anechoic data—smooth and flattish on and off axis seeming to be the desired pattern, combined with low-bass extension.

2.4 LEARNING ABOUT THE LOUDSPEAKER/ROOM SYSTEM

We listen in rooms, and while it is important to understand sounds that leave loudspeakers, it may be more relevant to understand the sounds that reach the listener. Figure 2.6 shows my first attempt, in 1985, at making sense of the complicated loudspeaker/room system. The idea was to use the NRCC listening room as the example room and try to predict the sounds arriving at a listener. A geometrical analysis of a loudspeaker and listener placed in that room identified the off-axis angles associated with floor, ceiling and side wall reflections and the path lengths to the listener. Perfect specular reflection was assumed at the reflection points, and the inverse-square law was used to calculate the propagation attenuation for each reflected sound. In reality, of course, there are many more reflections, but they would be lower in amplitude because they would have traveled farther and have experienced attenuation at multiple reflection points. The combined contributions of these other reflections were represented by a rough estimation of total sound power calculation based on 360° horizontal and vertical orbit measurements, modified according to the frequency-dependent absorption in the room as revealed in reverberation time (RT) measurements. There were assumptions and simplifications involved with the exercise, but this was to be a “first approximation” experiment to see what could be learned. Anechoic measurements were made that represented each major early reflection, appropriate attenuations were applied, the energies were summed and the results plotted. The computer-based measurement system made this possible.

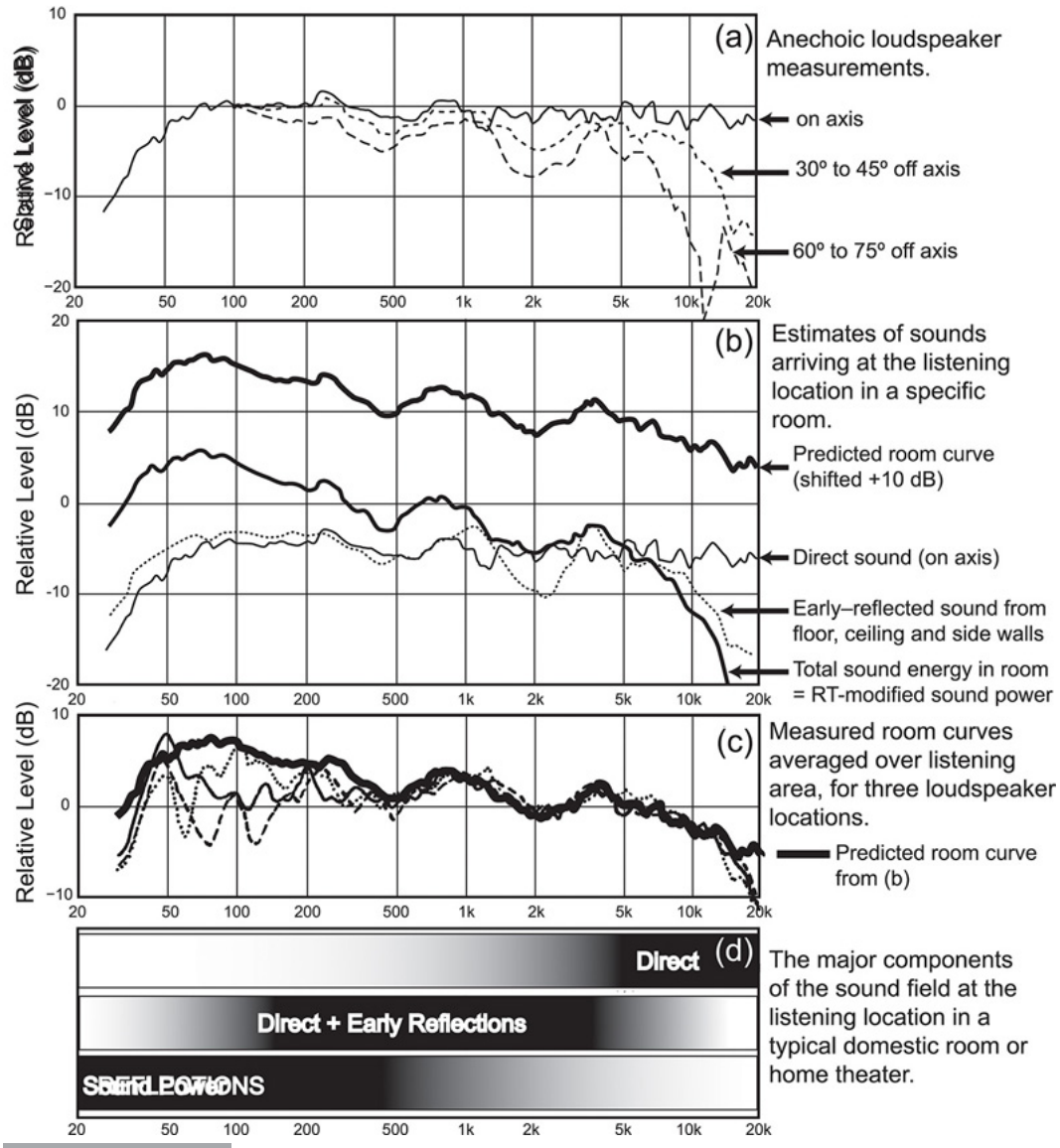


FIGURE 2.6 *A prediction of a room curve from anechoic measurements and a generalized description of the sound field at the listening location for a KEF 105.2 loudspeaker. Descriptions are in the text. Data from Toole (1986, part 2) as adapted for Toole (2015a, figure 4).* ⏎

Figure 2.6(a) shows anechoic loudspeaker measurements for the KEF 105.2 loudspeaker in which an excellent flat and smooth on-axis performance is evident. The irregular off-axis performance is also obvious.

Similar undulations can be anticipated in the total sound power output, which is dominated by off-axis radiated energy.

Figure 2.6(b) shows the principal components of the sound fields as they were estimated to be at the prime listening location in the NRCC (IEC 268-13 1985) listening room. In addition to the direct sound (the on-axis response) there is an energy sum of first reflections from the floor, ceiling and side walls derived from the appropriate anechoic off-axis responses, assuming spectrally perfect reflections and including inverse-square law attenuation. There is also an estimate of the contribution made by the total radiated sound power, calculated from a spherical area-weighted sum of measurements made on 360° horizontal and vertical orbits around the loudspeaker, modified according to the frequency dependent absorption in the room revealed in reverberation-time measurements.

The first observation is that if one wishes to anticipate how this loudspeaker might sound in a room, it is necessary to pay attention to sound power at low frequencies—it is the highest curve in Figure 2.6(b). The direct sound is dominant at the highest frequencies for the same reason. In the middle frequencies, all three components contribute significantly, so all three need to be measured. The on-axis curve by itself is insufficient data. Full 360° data, appropriately processed, is important information. An energy sum of the measurements contributing to these curves yields an estimate of a steady-state room curve, shown moved upwards by 10 dB for clarity.

The loudspeaker was then placed at three possible stereo-left locations in the listening room, and averaged measurements at six seats yielded the curves shown in Figure 2.6(c) with the prediction from Figure 2.6(b) superimposed. Above about 400 Hz the curves are essentially identical, and the prediction aligns well with the measurements. Below this the effects of standing waves in the room dominate, and the locations of the loudspeaker and the listener/microphone determine the acoustical coupling at different frequencies. The predicted curve provides an estimate of the upper limit of the steady-state sound levels, but destructive interference in the standing waves substantially reduces overall bass energy. Below about 400 Hz the

room is the dominant factor; above it, the loudspeaker is substantially in control. The differentiation between the frequency regions above and below about 400 Hz will be discussed in the next section, but it is clearly associated with the well-known *Schroeder frequency*, which in small rooms is also called the *transition frequency*.

Figure 2.6(d) attempts to illustrate the sounds arriving at the listening location in a typical domestic listening room or home theater—different portions of the radiated sound dominate at different frequencies, determined by the frequency-dependent directivity of the loudspeaker and the reflective nature of the room. Obviously, the frequencies at which the transitions occur will change with different loudspeaker designs and with different room acoustical configurations. In general, as loudspeakers become more directional and/or rooms become less reflective, the transitions move down in frequency.

These data illustrate some fundamentally important concepts:

- With sufficient anechoic data on a loudspeaker, it is possible to predict with reasonable precision mid- and high-frequency acoustical events in a listening space with known properties.
- There is an enormous difference between the spectrum of the direct sound arriving at a listener and that of the steady-state sound level that is achieved after reflected sounds arrive. The shape of a steady-state room curve is determined by the sound radiated by the loudspeaker modified by the geometry and frequency-dependent reflectivity of the room. In this example the early reflections were substantial influences on the shape of the room curve at middle and high frequencies. Frequency-dependent loudspeaker directivity—off-axis frequency response—is therefore a significant factor in what we hear in rooms.
- As reflections arrive, the room curve rises towards the predicted room curve, as the off-axis sounds add to the result. The overall downward tilt of the room curve is clearly related to the frequency-dependent directivity of forward-firing loudspeakers—woofers rapidly approach

omnidirectionality below about 200 Hz. With the desirable flat on-axis (direct sound) frequency response, there is progressively less energy radiated as frequencies rise. At very high frequencies the direct sound becomes progressively dominant.

- With no knowledge of the loudspeaker and no knowledge of the room acoustical properties, a steady-state room curve conveys ambiguous information and therefore cannot be used as a reliable "target" for performance evaluation—except at low frequencies, where loudspeaker directivity is not an issue.
- In normal rooms the on-axis frequency response is not the dominant physical factor. However, the on-axis/direct sound has a high priority in perception, establishing a reference to which later arrivals are compared in determining such important perceptions as precedence effect (localization), spatial effects and timbre. In this example, the poor off-axis performance dominated the in-room measurements and in listening tests caused audible timbral degradation. Equalization of the room curve will degrade the best and most important performance attribute of the loudspeaker: the on-axis/direct sound response.

> Equalization cannot change loudspeaker directivity; if it is the problem, the remedy is a better loudspeaker. Adequate anechoic data on the loudspeaker would have revealed the problem in advance of in-room measurements or listening.

- Below the transition/Schroeder frequency the room resonances and the associated standing waves are the dominant factors in what is measured and heard. The adjacent boundary effects are another contributing factor. These are unique to each room and are strongly location dependent. Only on-site measurements can reveal what is happening, and different loudspeaker and listener locations will result

in different bass sound quality and quantity. Seat-to-seat variations are a problem for multiple listeners.

> At low frequencies, steady-state measurements are trustworthy because most woofers are essentially omnidirectional. Equalization at low frequencies can reliably attenuate bothersome room resonances for a single listener. Satisfying multiple listeners requires different approaches, as will be discussed in Chapter 14.

- In domestic listening rooms and home theaters, or in any acoustically well-damped room like a cinema, traditional diffuse reverberation is significantly absent. As discussed in earlier editions of this book and in Toole (2006) and Toole (2015a), the sound fields in low-reverberation rooms are different from those described in classical acoustics. In addition to the effects of total absorption in the room, which is different at different frequencies, time-domain behavior at low frequencies is influenced by ringing of the obvious room resonances seen in the room curves in Figure 2.6(c). These modes need to be attenuated in any small room listening situation if seat-to-seat bass variations are to be reduced and timbrally neutral sound reproduction is to be achieved. The colloquial description “deep, tight bass” is a desirable goal.

This was an extremely useful exercise, yielding important insight into loudspeaker/room interactions. The research described above, and more, was described in two large papers in the *Journal of the Audio Engineering Society* (Toole, 1985, 1986) and the research became known worldwide. It was a foundation upon which to build.

With this evidence it was a short logical step to what has now come to be known as the spinorama, but it took several years for it to come to fruition. The intervening years were spent improving methods for accumulating

subjective and objective data, gaining confidence, and ultimately proving the value of the measured data as a basis for predicting sound quality.

2.5 A VIEW OF THE LOUDSPEAKER INDUSTRY IN THE 1980S

The accumulating archive of experimental results provided an interesting perspective on loudspeakers in the marketplace at the time.

Figure 2.7 is provocative in that it shows inexpensive loudspeakers being competitive with much more expensive ones in terms of sound quality. At moderate sound levels that can be true, especially if one is willing to accept limited bass extension. But most of the small inexpensive products misbehave at high sound levels, and almost certainly they do not have the compelling industrial design and exotic finishes that more expensive products have. It is possible to see a slight upward step in the ratings around $1000 ($2750) per pair, when the price could justify a woofer in a floor-standing product which extended the all-important bass. The dedicated professional monitor loudspeakers in these tests had problems. Professional systems were often equalized, which would alter the scores, but some of these loudspeakers had resonance and/or directivity flaws that were not correctable.

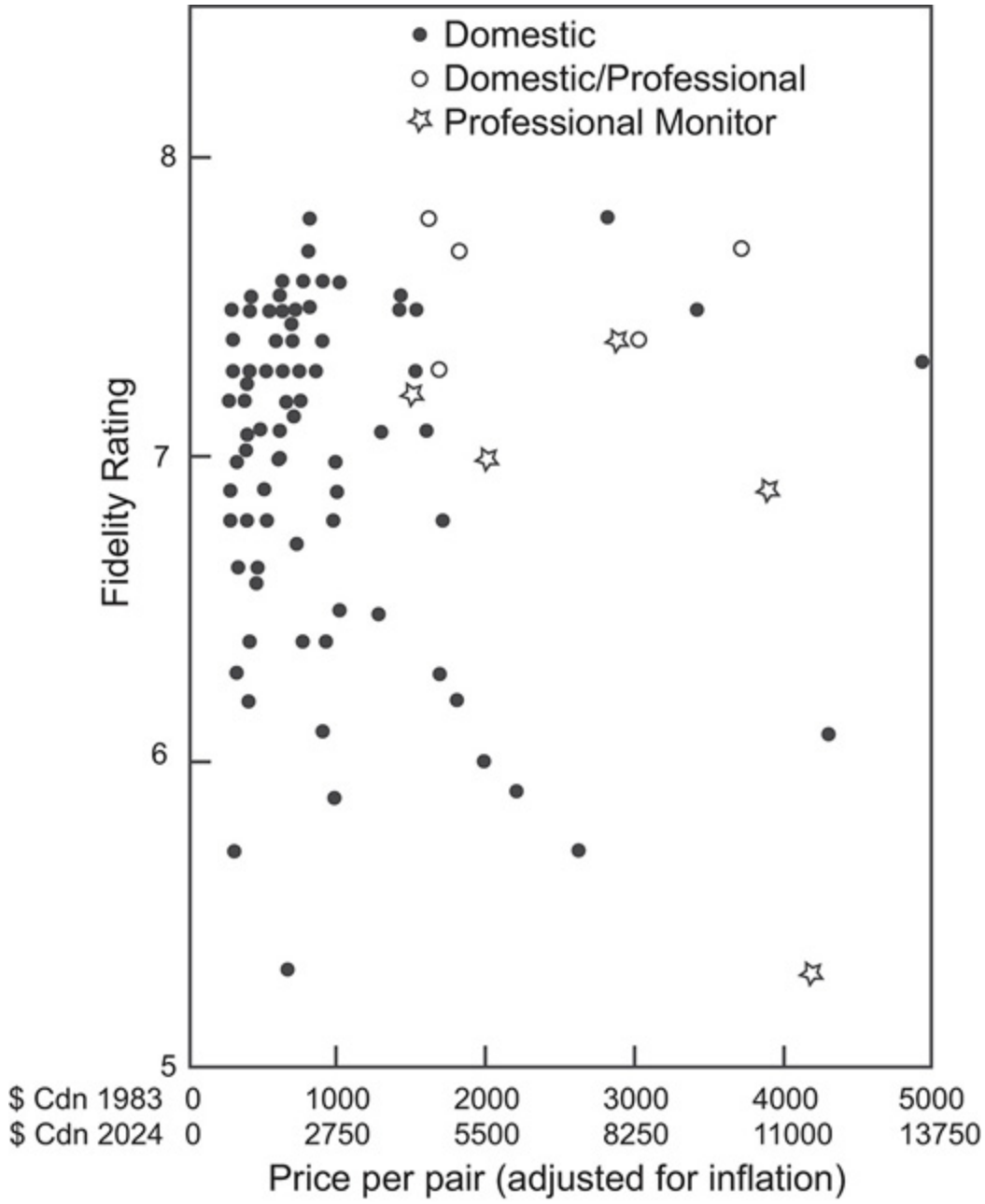


FIGURE 2.7 *Subjective "fidelity" ratings of loudspeakers evaluated in double-blind listening tests at the NRCC over the years 1980 to 1985. Prices have been adjusted for inflation for 1983. These were evaluated in groups of four, randomly selected from a larger number that included "anchor" products representing low scores and high scores. Each product was heard several times in the randomized groups, and these are the averaged ratings for all listeners. The "Domestic/Professional"*

loudspeakers were used in both applications. See Toole (1982, 1985 and 1986) for more details. ⏎

Historically, professional audio loudspeakers mainly addressed large audiences, with smaller versions in recording control rooms as monitors. Playing loud and not breaking were design objectives, with timbral accuracy a bonus. In contrast, consumer loudspeakers tend to be designed with timbral accuracy as the objective, and playing loud is a bonus. A few products bridge the gap, serving both markets. Over the years everything has improved. Current scientific knowledge combined with modern measurements and engineering methodologies yields an increasing proportion of products in the consumer-friendly upper left-hand region—high sound quality at moderate prices. Some professional monitor loudspeakers are now interchangeable with home audio applications.

2.5.1 International Interactions

Despite cost and complications in getting products through Canadian and US customs, a few US brands rented the NRCC facility for prototype development. Engineers from two large Japanese companies visited with measuring tapes and cameras, intending to duplicate the facility. Visitors from McIntosh were sufficiently impressed that they modified their approach to measurements and installed an anechoic chamber (www.roger-russell.com/lsd2.htm, look at 1977 [accessed July 16, 2024]). John Bowers of Bowers & Wilkins sent products from the United Kingdom to be tested. I also exchanged measurement data and visits with KEF, continuing lasting friendships with both Raymond Cooke, the owner, and Laurie Fincham, the engineering head. Kevin Voecks used the facility when he was with Mirage, and later when he was at Snell Acoustics, traveling from California to do so. Tomlinson Holman of THX used measurements in the NRCC lab to define some of the THX performance requirements and for a period had loudspeakers submitted for THX approval tested at the NRCC. John Eargle from JBL Professional visited the lab and, as a result, arranged for numerous products to be shipped for measurement. Harman International at

that time lacked the capability. It is possible that this encounter was a factor in my eventual invitation to join Harman.

With a limited travel budget, I was able to visit only a few loudspeaker manufacturers in Japan and England. I was impressed with some of the acoustical facilities in Japan, and some innovative and good-sounding loudspeakers came out of them. I visited the BBC Research Department in England, which had a distinguished history of loudspeaker research but, by then, it had all but ceased. My tour guide was Dudley Harwood, who had designed several BBC monitors for internal use in their often small, reflection-controlled broadcast studios, several of which I experienced. These loudspeakers were thoughtful designs for the purpose, and they influenced several designs of consumer products, especially in Britain. Measurements made at the NRCC indicated that there were quality control problems with some of their licensed builders—not all BBC monitors in circulation met the design specifications. The LS5/8 submitted for the CBC monitor tests was one, which led to its rejection. The widely praised LS3/5A was experienced in multiple quite different sounding and measuring versions from supposedly licensed manufacturers. It seems that these companies were simply exploiting the BBC reputation. Manufacturing variation in transducers is an industry problem that persists today, but some of these appeared to be flawed designs.

2.6 TURNING POINTS IN RESEARCH AND CAREERS

As happens from time to time, politicians change the emphasis on how taxpayers' money is spent, and the NRCC is a government organization. At a point, scientists were encouraged to seek funding from the industrial sector to help support their research. Fortunately, by then the loudspeaker research had strong connections to several Canadian audio companies, a few of which were doing quite well. I arranged a meeting of manufacturers to consider a proposal: a government-subsidized collaboration between a non-profit Canadian Audio Research Consortium of manufacturers and my NRCC research group. The proposal was accepted, and the Athena project

was launched. It was the first such alliance between the NRCC and industry. The work focused on low-frequency sound reproduction in small rooms, the one aspect of sound reproduction that loudspeaker manufacturers cannot control without electronic equalization capability. A suite of dimensionally adjustable listening rooms was built to explore room effects, and specialists were hired to address the digital electronic aspect of the project. Research was well under way when, in 1991, my phone rang.

The caller was a headhunter looking for a corporate vice president of engineering for Harman International Industries, Inc., in Northridge, California. Harman was a large corporation, home to many well-known brands and active in all aspects of sound reproduction: consumer, professional, cinema, and automotive, so there would be many opportunities and challenges. The headhunter had done his homework, and he laid out a candid description of the corporate situation and how I might fit into it. I felt that much of the core research had been done and was enticed by the prospect of participating in its application in new domains. My job "interview" was a 4-hour lecture to the assembled Harman engineering staff and management—I had just completed a lecture tour of several AES chapters and was prepared. I accepted the job, with the title of corporate vice president of acoustical engineering. As a corporate officer I had no staff, no line authority, but I could interact with any of the brands under the Harman umbrella, comprising the well-known consumer brands of JBL, Infinity, Harman Kardon, Lexicon, Mark Levinson, and soon Revel, along with the several professional audio and automotive audio brands and activities. It was an interesting position with both advantages and frustrations.

My advice led to measurement equipment and anechoic chambers being installed to improve design capabilities, but subjective evaluations were needed to confirm success. It was not long before I suggested that the company needed its own research group to continue the work begun at the NRCC—its output would be knowledge, not products. It would also create a listening facility and administer the double-blind listening tests necessary to confirm the performance of loudspeakers as they were developed.

Remarkably, Chairman Dr. Sidney Harman and senior management agreed, and more importantly, they agreed that the new corporate researchers could publish their results, following the scientific tradition of open information exchange. This of course informed the entire audio industry, not just Harman. It was an exceptionally generous decision as many corporations insist on keeping such knowledge as their own, although it usually leaks out as employees migrate to other companies. The future could not have been predicted, but in retrospect it was an astute decision because the name Harman has become known and respected for its contributions to the science of audio as well as for its products.

Some cynics still see this as a marketing exercise, but they are wrong. Those individuals do not realize that the bulk of the loudspeaker research was done over a 25-year period in Canada, at the National Research Council, a taxpayer-funded research establishment. The published research findings benefited many manufacturers. It was because of that successful research that I was offered the job at Harman.

The fact is that within the company the scientific guidance was not immediately accepted by all design engineers. This was especially true in foreign countries as Harman International expanded its operations in Europe and Asia. Consequently, delivering educational tutorials became a continuing component of my task, including delivering numerous Audio Engineering Society lectures and dozens of CEDIA (Custom Electronics Design and Installation Association) courses to inform consultants and installers of audio equipment. Now, years later, the scientific knowledge gained during the NRCC and Harman years is being used by researchers and loudspeaker engineers throughout the world. Some of Harman's competitors were responsible for creating the ANSI/CTA-2034-A (2015) loudspeaker measurement standard that incorporates results from the NRCC and Harman research.

2.7 THE "SPINORAMA" EMERGES

The next evolutionary step in loudspeaker measurements took place after I moved to Harman, where there were new facilities and a new research group. Faster measurement technologies meant that more data could be acquired in a reasonable time, thus the angular resolution was changed to 10° for both rotations. Seventy measurements described the frequency response of a loudspeaker. The new version of the two-axis measurement scheme is shown in Figure 2.8.

The low-frequency chamber calibrations were referenced to both ground plane and tower measurements and are, as before, most accurate with closed-box woofer systems. The discussion following Figure 2.3 applies here as well.

After the data are accumulated, the next step is to process it for visual display. Figure 2.9 illustrates the basic method. The measured product is an Infinity Prelude MTS (ca. 2000), a hybrid design with an active woofer that reliably achieved high subjective ratings in double-blind evaluations. Its extended bass was a contributing factor. The family of curves in Figure 2.9(d) is the full spinorama—a name that has no specific origin, but which is obviously associated with the fact that the loudspeaker is rotated, "spun", to collect the basic data set. However the term originated, it has become the default descriptor.

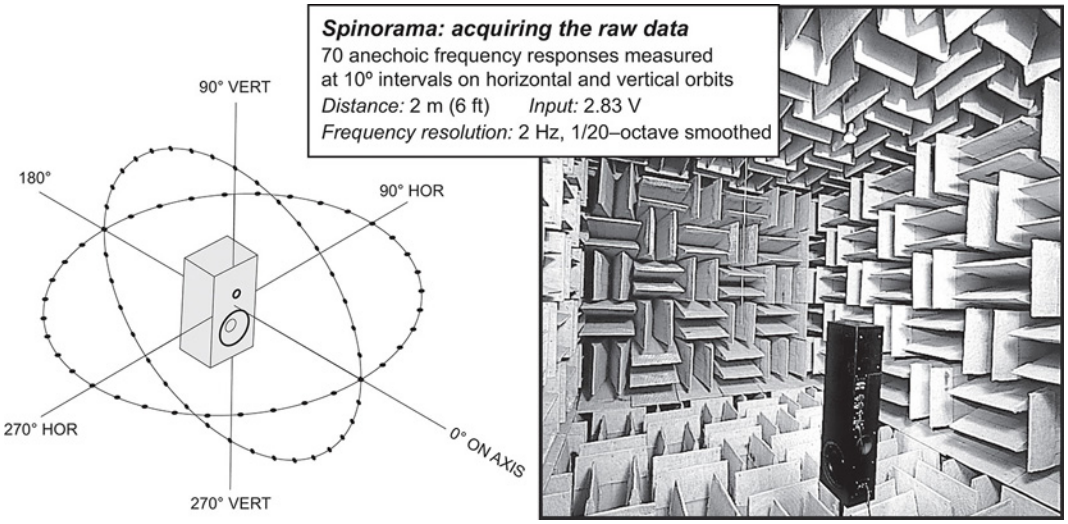


FIGURE 2.8 *The process used in acquiring the raw data used in the spinorama. The chamber shown has 4 ft (1.2 m) wedges and*

anechoic ±0.5 dB at 1/20 octave from 60 Hz to beyond 20 kHz, and has been calibrated to be ±0.5 dB from 20 Hz to 60 Hz at 1/10-octave resolution for the reference microphone and loudspeaker locations. Measurements are made at 10° increments on horizontal and vertical orbits, 70 frequency responses in all, at a distance of 2 m (6 ft) with an input of 2.83 volts—the sound level being corrected to the standard 1 m distance for sensitivity evaluation. The frequency resolution was 2 Hz, smoothed to 1/20 octave. ⏎

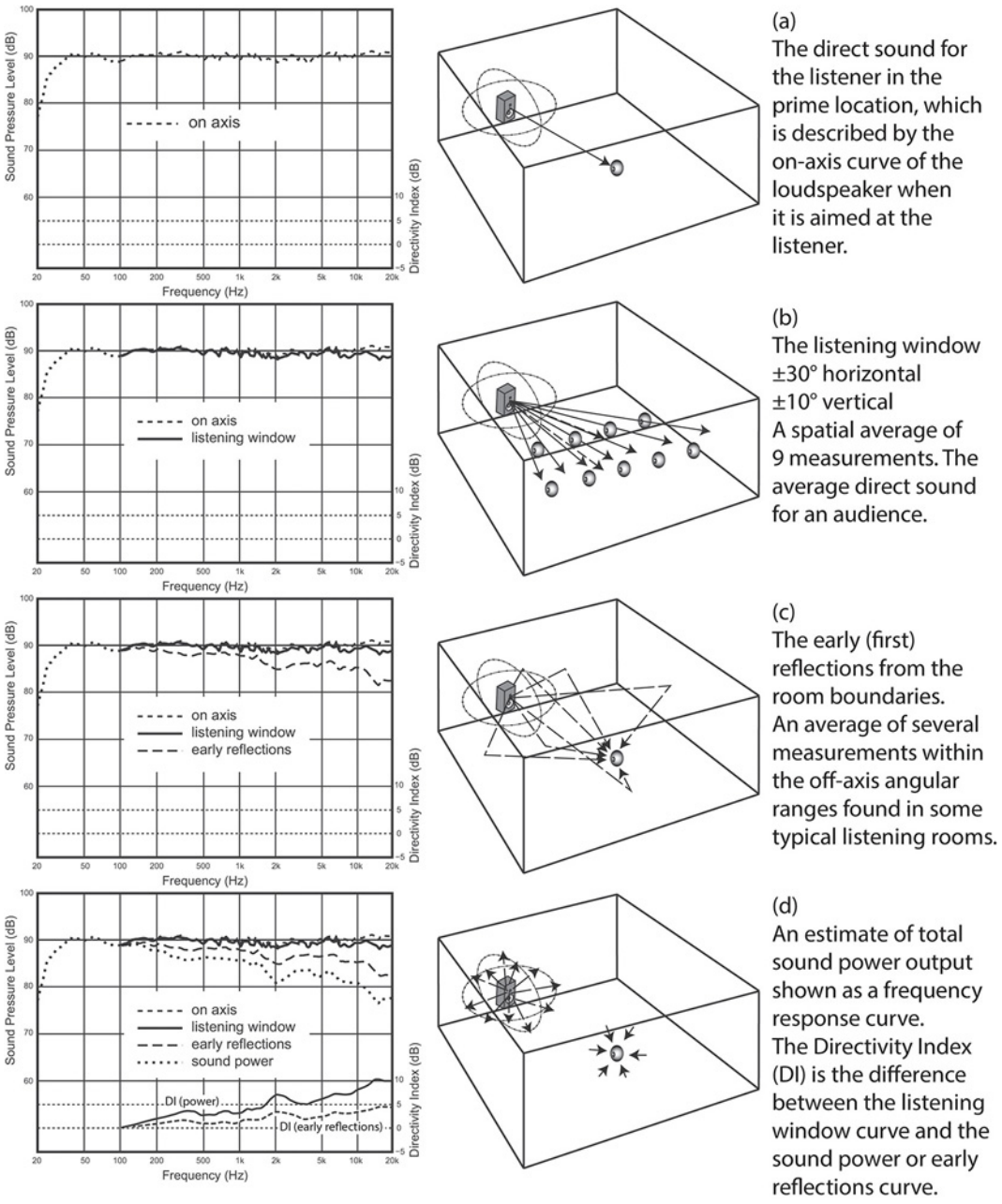
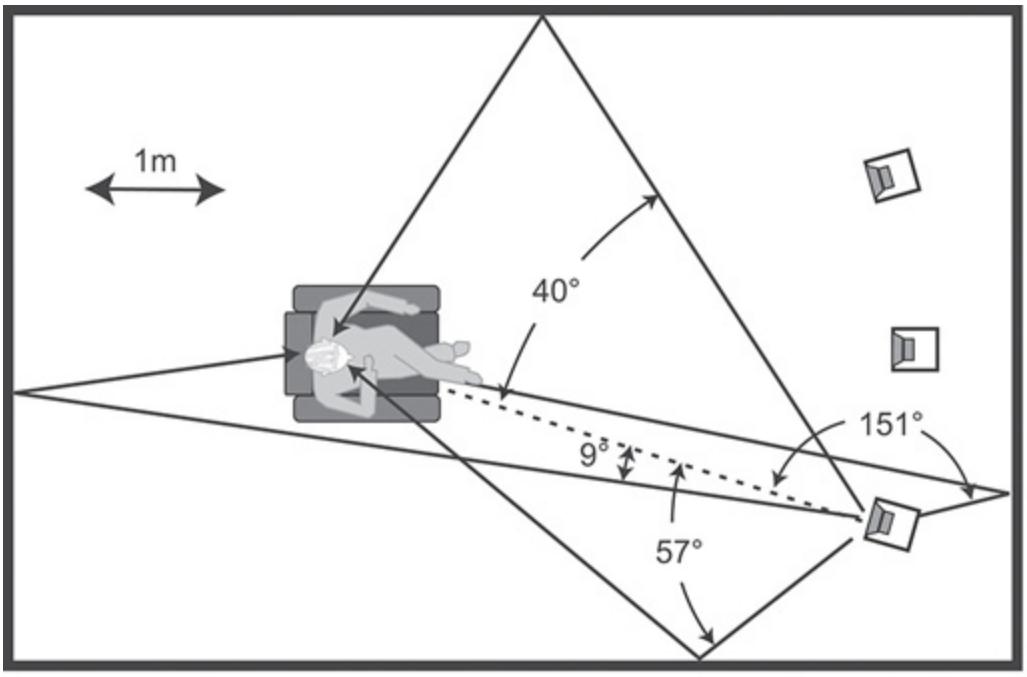
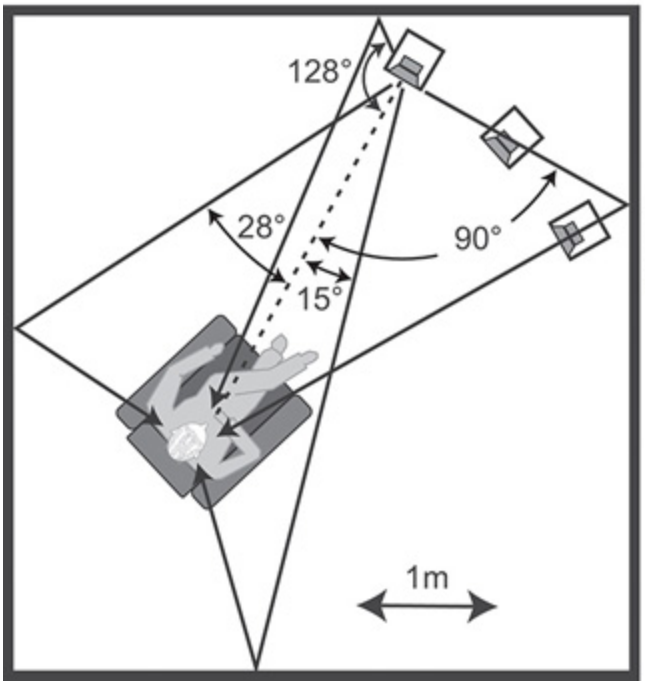


FIGURE 2.9 *A pictorial description showing loudspeaker and sound rays radiating to a listener in a room for each of the individual curves post-processed from the data set. The direct sound, listening window, early reflections and sound power curves are shown as the full spinorama develops and are shown in (d).*

The curves shown are:

- The *on-axis frequency response* is the universal starting point for loudspeaker evaluation, and, in many situations, it is a fair representation of the first sound to arrive. However, as shown in the Devantier (2002) survey, over half of those investigated had the prime listening position 10° to 20° off axis, providing a justification for the following measure: a version of the listening window used in the NRCC data.
- The *listening window* is a spatial average of the nine frequency responses in the ±10° vertical and ±30° horizontal angular range. This embraces those listeners who sit within a typical home theater audience, as well as those who disregard the normal rules when listening alone. Because it is a spatial average, this curve attenuates small fluctuations caused by acoustical interference, something far more offensive to the eye than to the ear, and reveals evidence of resonances, something the ear is very sensitive to. Interference effects change with microphone position and are attenuated by the spatial averaging, while resonances tend to radiate similarly over large angular ranges and remain after averaging. Bumps in spatially averaged curves tend to be caused by resonances.
- The *early reflections* curve is an estimate of all single-bounce first reflections in a typical listening room. Measurements were made of early reflection "rays" in 15 domestic listening rooms (Devantier, 2002). Figure 2.10 shows ray diagrams of some possible horizontal plane reflections.





Angles of early-reflected sounds relative to the on-axis direct sound.

FIGURE 2.10 *Examples of horizontal-plane early (first) reflections in listening rooms arranged with the primary listening axis along a parallel room axis and also on an oblique room axis. The angles shown are relative to the direct sound axis shown as a dashed line.* ⏎

Figure 2.10 shows that in the parallel arrangement the first lateral reflections are associated with moderately large off-axis angles: 40° and 57°, angles at which there will be significant mid- and high-frequency energy radiated by most loudspeakers. The sound reflected from the wall behind the loudspeaker will have little mid-high frequency energy from forward-firing loudspeakers. The sound from the wall behind the listener is close to the on-axis (i.e. direct) sound and will be a significant factor. Effective sound absorbing material on the wall behind the listener is likely to be advantageous; indeed, if the separation is small, it is necessary (see Section 13.6).

In the oblique arrangement the first "lateral" reflections are associated with very large off-axis angles: 90° and 128°, where mid and high frequencies are greatly attenuated. The "rear" reflections fall within the listening window angles (strong mid and high frequencies). In small rooms the reflecting surfaces might benefit from absorption or diffusion. In a large room, with the walls behind the listener at a greater distance, the listener is in a predominantly direct sound field at mid-high frequencies. It is a good option, albeit not a popular one, perhaps for aesthetic reasons. Two of my personal listening rooms have used it.

A survey of typical domestic listening setups resulted in 70 measurements of this kind, leading to an estimate of the first reflections arriving at the listening location in an "average" room (Devantier, 2002). It is the average of the following:

- Floor bounce: average of 20°, 30°, 40° down
- Ceiling bounce: average of 40°, 50°, 60° up
- Side wall bounces: average of ±40°, ±50°, ±60°, ±70°, ±80° horizontal
- Front wall bounce: average of 0°, ±10°, ±20°, ±30° horizontal
- Rear wall bounces: average of 180°, ±90° horizontal

The summary illustration for the important forward hemisphere of a forward-firing loudspeaker is shown in Figure 2.11.

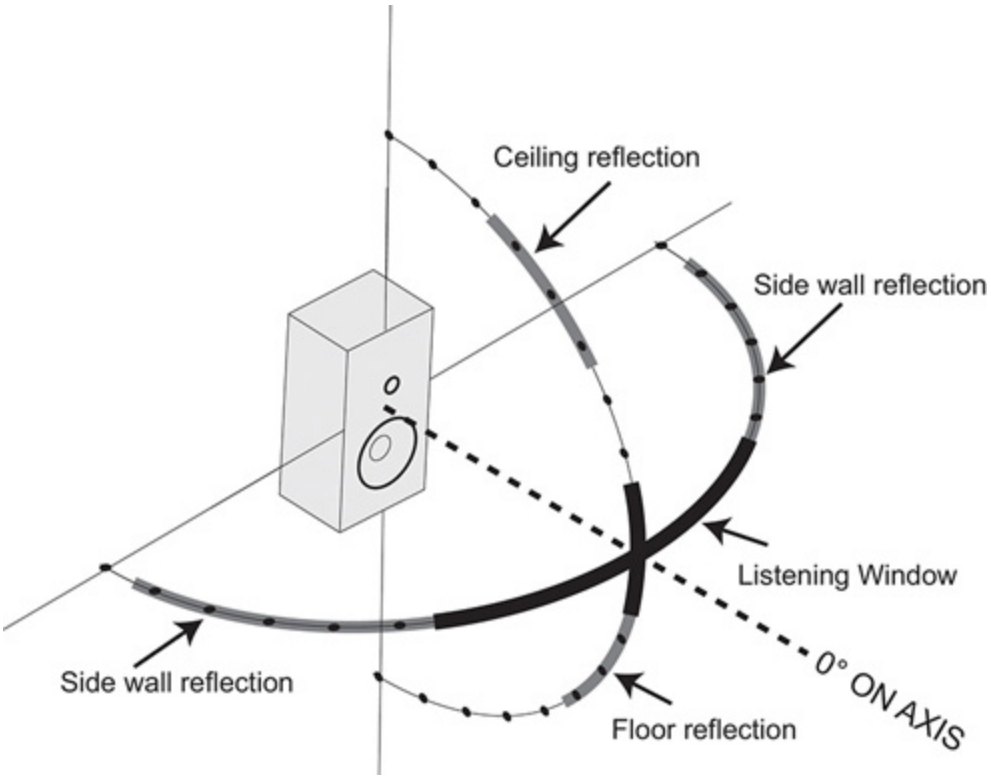


FIGURE 2.11 *A visual portrayal of the angular ranges considered in calculating the listening window and early reflections in the front hemisphere of a forward-firing loudspeaker. The early-reflections calculation includes the horizontal portion of the listening window.*

The number of averages mentioned in that description may make it seem as though anything useful would be lost in statistics. However, this turns out to be a very useful metric for two reasons. First, it provides an estimate of the contribution to direct and early-reflected sounds arriving at a listener in a typical room. These are very influential in establishing timbral and spatial qualities. Second, being a substantial spatial average, a bump that appears in this curve, as well as in other curves is strong evidence of a resonance.

- *Sound power* is a measure of the total acoustical energy radiating through an imaginary spherical surface surrounding the loudspeaker. In the present context it is something of limited relevance because all of the sound radiated by a loudspeaker is not equally important to what is

measured and heard at a listening position in a rectangular semi-reflective room: a normal listening room. Ideally, such a measurement would be made at many closely spaced points over the entire surface of the sphere. In the spinorama the sound power is estimated by calculating an energy sum of the 70 measurements on two circular orbits, with individual measurements weighted according to the portion of the spherical surface that they represent. The weighting coefficients are in ANSI/CEA-2034-A (2015). Thus, the on-axis curve has very low weighting because it is in the middle of other, closely adjacent measurement points (see the perspective sketch in Figure 2.8), and measurements farther off axis have higher weighting because of the larger spherical surface area that is represented by each of those measurements (the area between the orbits). Because it is an energy sum, the final curve must be computed from sound pressures, converted from dB SPL, which are squared, weighted, summed and then converted back to dB. The result could be expressed in acoustic watts, the true measure of sound power, but here is left as a frequency-response curve having the same shape and normalized to the other curves at low frequencies. This serves the present purposes more directly. Any bump that shows up in the other curves, and persists through to this ultimate spatial average, is a notable resonance.

- *Directivity index (DI)*: Traditional DI is defined as the difference between the on-axis curve and the normalized sound power curve. It is thus a measure of the degree of forward bias—directivity—in the sound radiated by the loudspeaker. It was decided to depart from this convention because it is often found that, because of symmetry in the layout of transducers on baffles, the on-axis frequency response contains acoustical interference artifacts, due to diffraction, that do not appear in any other measurement. It seems fundamentally wrong to burden the directivity index with irregularities that can have no consequential effects in real listening circumstances, or to spectrally smooth the data to disguise the problem (e.g. using 1/3 octaves).

Therefore, the spinorama DI has been redefined as the difference between the listening window curve and the sound power. In most loudspeakers the difference is small; in highly directional systems it can be significant. In any event, for the curious, the raw evidence for the classic DI is there to inspect. A DI of 0 dB indicates omnidirectional radiation. The larger the DI, the more directional the loudspeaker in the direction of the reference axis. This is illustrated in Figure 2.12. Because of the special importance of early reflections in what is measured and heard in rooms a second DI is calculated, the Early Reflections DI, which is the difference between the listening window curve and the early reflections curve—this is possibly the more important measure.

2.8 INTRODUCING THE RESEARCH TEAM

Over the years I have had the pleasure of working with many talented people, from technicians through to research scientists. Among them, the following three individuals have had long-term relationships with me and have made significant contributions to the science and technology of sound reproduction.

2.8.1 Sean Olive

Born in Brockville, Ontario, Canada, in 1959, Sean set off to be a pianist and graduated from University of Toronto in 1982 with a bachelor's degree in music. Deciding that recording music might have a better future than creating it, he modified direction to sound recording. He joined the Tonmeister (sound recording) program at McGill University in Montreal. While there he engineered and produced numerous concert performances and studio sessions.

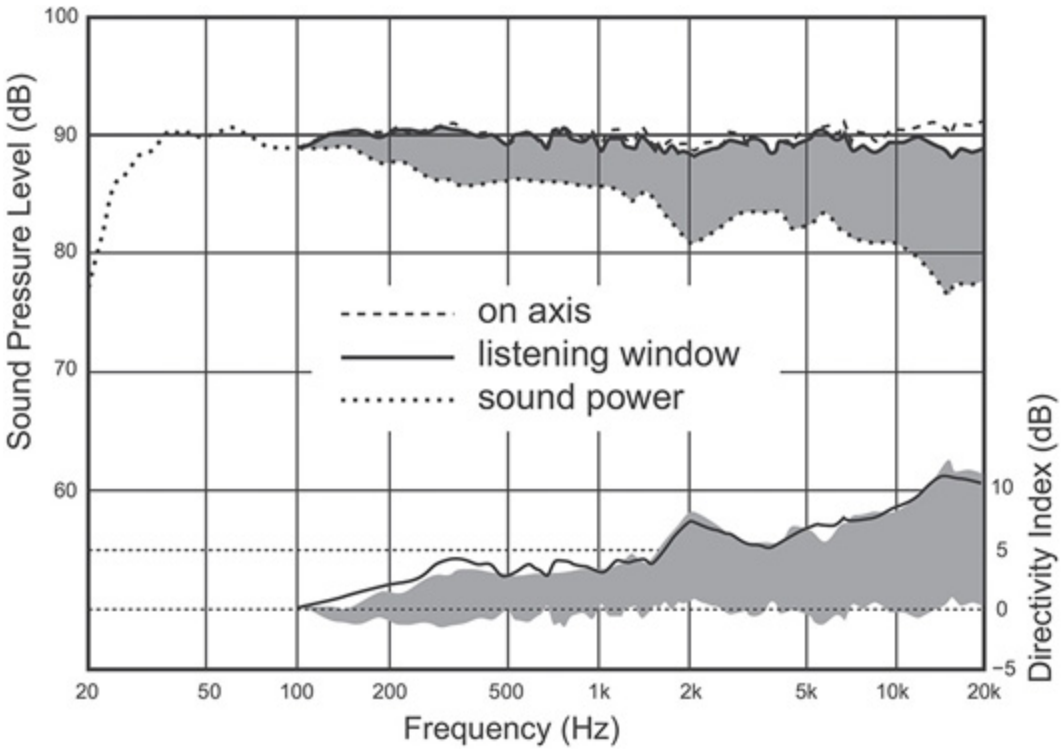


FIGURE 2.12 *An explanation of the spinorama directivity index (DI) based on the difference between the listening window curve and the sound power curve. Also shown is the difference area inverted and superimposed on the DI, demonstrating that the inverted DI is equivalent to the sound power of this loudspeaker if it were equalized to be flat in the listening window. This can be visualized by imagining that the bottom of the shaded area in the lower illustration is adjusted to be flat and smooth—the top of the shaded area would then align with the DI curve. The same process can be applied to the Early Reflections DI.*

In 1985 Sean visited the NRCC on a McGill student tour, getting a brief lecture on the research as it was, touring the facility and taking part in a demonstration listening test. He was intrigued by the connection between what was heard and what was measured. To this potential recording engineer, it was a challenge to some of the current thinking in professional audio. Different recording engineers had different preferences in loudspeakers, and at the time, some were significantly colored. There was

no assurance that the art created in the recording control room would ever be heard again. He became a guest researcher at the NRCC, supported by some Canadian loudspeaker companies, including providing THX with measurements on loudspeakers submitted for their approval. Upon receiving his master's degree in music (sound recording) in 1986, a position was created allowing him to become a permanent staff member.

The extra pair of hands and thoughtful collaboration were welcomed. The two of us shared in investigations focused on the audibility of resonances in loudspeakers, the audibility of early reflections in listening rooms, and the performance of recording microphones. In 1993 I hired Sean as the first member of the new Harman Corporate Research Group. He was pleased to come to sunny southern California. In addition to his continuing contributions to the research, Sean later returned to his alma mater, McGill University to earn a PhD. He recently changed focus to headphone sound quality among other things, including contributing to this book. Sean is continuing his research as an independent consultant.

2.8.2 Todd Welti

The second permanent member of the research group was Todd Welti, born in 1962 in Los Alamos, New Mexico. He received a B.S. degree in electrical engineering from Cogswell College North, Everett, Washington, in 1989, and in 2000 he received an international masters in acoustics degree from Aalborg University, Denmark. He was an experienced acoustical consultant, having spent 9 years with Michael R. Yantis Assoc. commissioning sound-reinforcement systems from performance halls to multipurpose arenas. He was also involved in room acoustics design for many lecture rooms, concert halls and other critical spaces. While this experience was valuable, it was his skills in mathematical modeling that were central to the development of multiple subwoofer methods as a means of diminishing the detrimental effects of small-room resonances, especially seat-to-seat variations in bass response. I had used a simple two-subwoofer room mode cancellation in a room in my home in Canada, and felt that a more general solution, applicable to many rooms was possible. Todd was

able to identify a collection of subwoofer arrangements that reduced seat-to-seat variations in rectangular rooms. Bothersome room modes are attenuated, and equalization works for several listeners, not just the one in the prime location.

Then, I challenged him to find a solution that would work in rooms of non-rectangular shapes—such as my own home theater. In this effort he was joined by Allan Devantier, to develop Sound Field Management a powerful multiple-subwoofer scheme that employed on-site measurements, an optimization algorithm and signal processing in the signals supplied to each subwoofer. The room shapes and the locations of the subwoofers were more flexible.

With both systems bass in small rooms was much improved. Multiple-subwoofer schemes have now been widely adopted in the industry with several commercial variations following the initial inspiration. Todd has gone on to develop auralization techniques and system equalization methods, as well as collaborating in headphone research. He is a contributor to this book.

2.8.3 Allan Devantier

Allan Devantier was born in 1964 in Newmarket, Ontario, Canada. He received a bachelor's degree in electrical engineering technology with honors in 1987 from Ryerson Polytechnical University in Toronto, Canada. Showing extraordinary enthusiasm, he was a private user of the NRCC loudspeaker measurement facility in his late teens. After graduation he worked for two Canadian loudspeaker manufacturers and was about to launch his own brand when I suggested that he apply for a job at Harman. He joined the Harman Consumer Group in 1993 and provided engineering leadership in both JBL and Infinity, where he was systems engineer on Infinity's excellent Prelude MTS loudspeaker system. He is the inventor of Infinity's Room Adaptive Bass Optimization System (RABOS) and coinventor of the Ceramic Metal Matrix Diagram cone material (CMMD) which continues in various forms. Along the way, Allan made extensive use of the research facility in his product development, collaborated in some of

the research and initiated the final form of the spinorama method of presenting anechoic loudspeaker measurements. He then joined the corporate research group, working with Todd Welti on the multiple subwoofer project, and leading the design of the JBL Professional M2, another benchmark loudspeaker. He was hired by Samsung in 2013 to set up a loudspeaker development facility and research group in California to address their TV-centric product needs, which is now a major category in the audio industry. This includes participating in the development of IAMF, an open-source three-dimensional spatial audio technology. In 2016 Samsung purchased Harman International, which brought everybody together under a new corporate umbrella, although Harman has continued to function with some independence. It is a small world.

2.9 MEASUREMENTS HAVE CHANGED THE LOUDSPEAKER INDUSTRY

Loudspeaker measurements have existed for a very long time. Most of them were not very revealing of loudspeaker performance for a variety of reasons, and a significant problem was that they were not standardized, so comparisons among products from different origins was challenging. The main problem was that few useful measurements made it into the public record. They stayed in the engineering lab, if they existed at all. Published frequency response curves were often artistic interpretations, with no information about how they were measured. Mostly, they were numerical specifications with overly generous tolerances, and again little or no information about their origins. Marketing departments didn’t want to be technically specific, so customers have been uninformed. They have had to rely on their own subjective evaluations if they were possible, or unreliable subjective reviews which in the internet age exist in abundance.

Now, it is very gratifying to see evidence of scientific research results being applied to loudspeakers by many manufacturers worldwide. The average sound quality has noticeably improved as a result, and several products are arguably as about good as they need to be, given the variability

in recordings and listening rooms. “Perfection” in loudspeakers may or may not be completely rewarding if achieved, but it is a worthy goal. If all loudspeakers were equally well designed, they would all sound remarkably similar, neutral. As things stand, the highest-rated loudspeakers in double-blind listening tests often end up in statistical ties—with small variations in opinions with different recordings.

> Loudspeakers are not the “weakest link” in the audio system. Low-frequency resonances in small listening rooms significantly degrade sound quality in all systems—differently.

Technical solutions to the remaining problems exist, so the weakest links are not as weak as they once were, and the future is looking better.

Competition in the marketplace requires product differentiation. There is a never-ending struggle to be distinctive in some way, talking points for marketing, hopefully without compromising sound quality. Concepts that are just differences are promoted as advantages. Fortunately, competent engineering can accommodate many visual forms, and transducer technologies, and still deliver excellent, basically neutral, sound. As expected, though, human nature remains, and there are still examples of aberrant designs based on personal opinions of people with “special” tastes, and others that follow misguided technical objectives. *Caveat emptor*.

Consumers, sadly, are not in a good situation to recognize excellence in sound quality. Not only are the circumstances for subjective evaluation not blind or double-blind, but most available methods for forming opinions are not advantageous; see Section 1.4. The traditional brick and mortar “audio shop” is disappearing, which adds to the challenge. For this reason, I believe it is timely that there are now technical methods of identifying loudspeakers with the potential of delivering good, neutral sound. Comprehensive anechoic measurements in a variety of presentation styles are helpful.

All trustworthy technical data on loudspeakers originate in anechoic chamber or equivalent signal-processed anechoic measurements such as widely available FFT time-windowed measurements or the Klippel near-field scanner (www.klippel.de/products/rd-system/modules/nfs-near-field-scanner.html [accessed August 2024]). One often can evaluate on- and off-axis performance in sufficient detail to get a reasonable impression of performance. My compliments to those who have the courage to publish credible data.

John Atkinson of www.stereophile.com has provided consistent loudspeaker measurements for many years. He employs a combination of methods to generate useful quasi-anechoic information on products that were reviewed. There are flaws, which he admits, but the accumulated data are revealing. Interestingly, they occasionally contradict the subjective reviewer's comments, almost always revealing the weakness in the "take-it-home-and-listen-to-it" subjective evaluations, not errors in the measurements.

Amir Majidimehr at www.audiosciencereview.com and Erin Hardison at www.erinsaudiocorner.com both have Klippel near-field scanners and publish accurate and revealing data on loudspeakers in spinorama form as well as other styles of presentation. James Larson at www.audioholics.com does comprehensive outdoor measurements that yield spinorama-formatted results, and more, which he combines with subjective evaluations. For published anechoic data in my original NRCC measurement format, www.soundstagenetwork.com shows data on many loudspeakers. These data both inform and moderate their subjective comments, in the absence of double-blind listening.

The website www.spinorama.org reveals links to spinorama measurements on hundreds of loudspeakers from many manufacturers and independent sources evaluated using the ANSI/CTA-2034-A (2015) standard method. In a few cases, it is impossible to learn much from the published spinorama data because the curves have been overly smoothed (a persistent marketing department practice) or lack adequate resolution at lower frequencies caused by the time-windowed FFT measurement method.

Both violate the frequency-resolution requirements of the standard, and such data should not be trusted.

These spinoramas and other published anechoic data on manufacturers' websites reveal numerous well-designed, relatively neutral loudspeakers in the marketplace, an encouraging fact. Nevertheless, not all products are flattered by the data. Negative reviews displease advertisers, which is a serious problem for publications. It is truly unfortunate that they need advertising revenue from purveyors of poorly designed products, exotic wires, overpriced audio "jewelry", and the like, that are of no audible benefit. Wealthy benefactors are needed to support competent independent evaluations. Subjective reviewing knows no boundaries—"it's my opinion, prove me wrong". They generate long essays on the "sound" of interconnects, power cables, anything. Everything yields "tighter bass", "more detail", "better imaging" and profit.

Single-number ratings of various aspects of performance have been derived from spinorama data by some users, and these are appealing. However, they are simplistic interpretations and can mislead. It is easy to find examples of loudspeakers with low numerical ratings because of one correctable fault, such as limited bandwidth (add a subwoofer), a broadband spectral tilt (use tone controls) or a strong resonance (compensate with parametric equalization). Anechoic data reveals the problems and suggests solutions.

It is worth learning how to interpret curves. It is clear from spinorama data whether a loudspeaker is likely to achieve very high or very low sound quality ratings. The difficulty is in deciding among the vast middle ground of loudspeakers that are neither the best nor the worst. They all have different kinds of imperfections, and individual listeners may express their own views on how they make their trade-offs, such as more extended bass vs. a noticeable mid-frequency coloration. The possibilities are countless.

The good news is that for those with some technical expertise, equalization *based on anechoic data* is an option that can elevate the performance of slightly flawed loudspeakers to a higher category. This is what is done in active loudspeakers.

Unfortunately, there is the longstanding industry problem of production variations to deal with. A design begins with a "golden prototype". In responsible manufacturers the design then moves through preproduction phases during which sample-to-sample variations in components are assessed and mass-production performance of the entire system is estimated. Samples are produced, and at some point, definitive measurements are made on a single representative sample, or an average made of a few samples. In an ideal situation, this is what the published data represents.

From there follows the task of duplicating this performance in the hundreds or thousands of replicas made to be sold. End-of-line testing remains a challenge, especially, but not exclusively, in products at affordable prices—it takes time and costs money, especially when deviating products must be repaired or scrapped. Some of what one pays for in premium loudspeakers *might be* extra attention to ensure that the customer gets a properly manufactured product.

It is interesting to contemplate the possibility that some listener dissatisfaction with stereo soundstage and imaging is the result of dissimilar left and right loudspeakers. Getting similar steady-state room curves is not the answer because sound localization (imaging) is determined by the direct sound, and room curves are dominated by reflected sounds.

Having widely available trustworthy measurements adds a whole new dimension to an industry that for too long has drifted along relying on inadequate measurements, unreliable subjective evaluations and published "specifications" that are often useless.

2.9.1 A Real-World Problem for Loudspeaker Users

Figure 2.13 illustrates an example of the enormous difference between typical specifications and useful measurements. An explanation of the data is included for educational purposes.

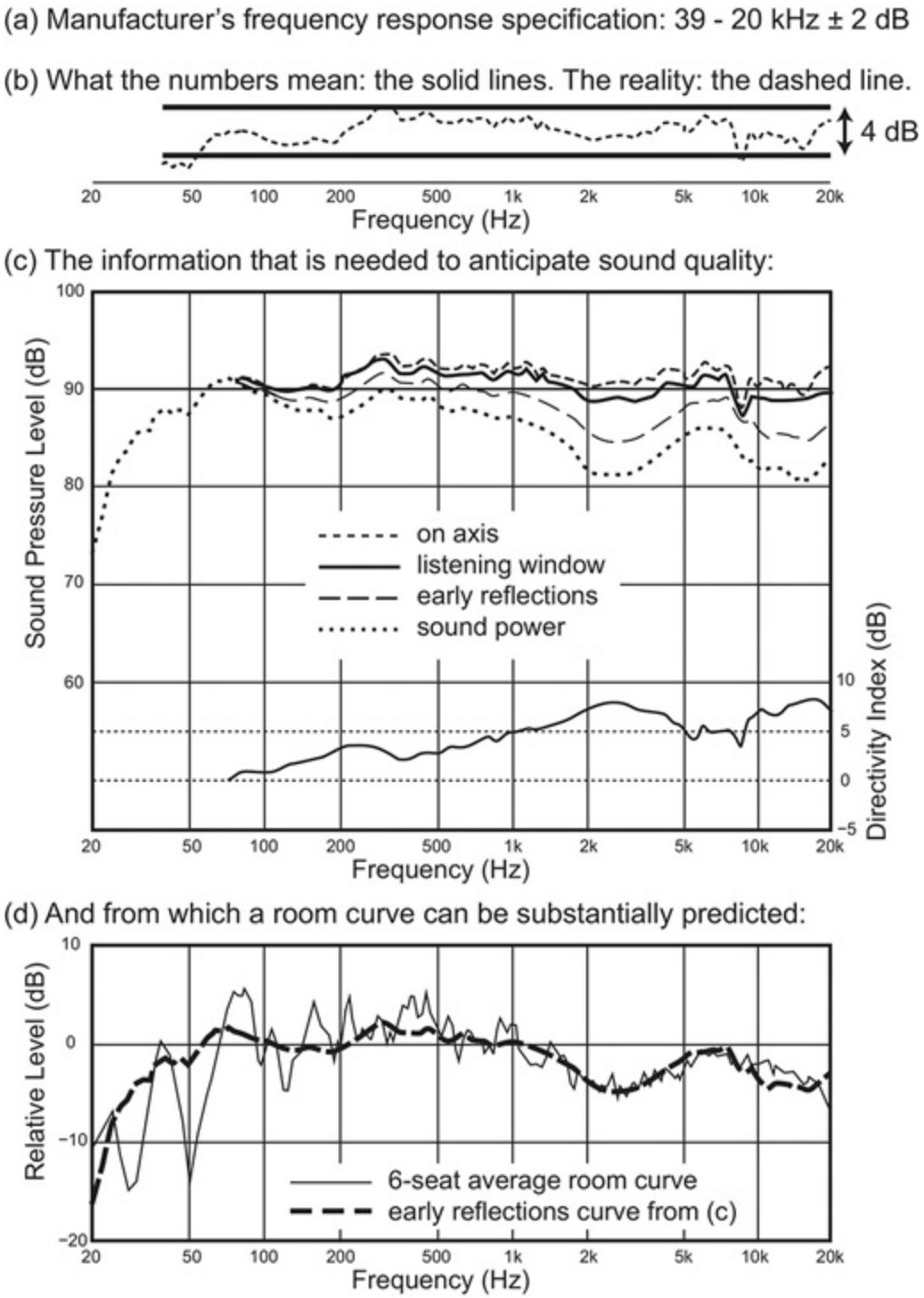


FIGURE 2.13 *The manufacturer's specification and measured data on a B&W 802N, introduced in 1999.*

The simple numerical data in Figure 2.13(a) is all the information a customer is given about the frequency response of this product: 39 Hz to 20 kHz ±2 dB. This specification is for the on-axis, direct sound performance only, although this fact is not often revealed. The bandwidth is adequate for most music, but movies and organ music would benefit from output at lower than 39 Hz. The loudspeaker might have it, but this information does not reveal it. We need to know the frequency at which the bass output is at −5 or −10 dB, not −2 dB, allowing for the acoustical gain in small listening rooms. See Figure 2.5.

As evidence of broadband output uniformity—smoothness—the ±2 dB tolerance is an improvement on the common ±3 dB, but it is bettered by a few that claim ±1.5 dB. What really matters, though, is what happens within those tolerance limits. Is it a straight line with smooth roll-offs at both ends (perfection) or is it something else?

Here the dashed line in Figure 2.13(b) tells us that it is something else. Broad low-Q undulations, with superimposed bumps and wiggles together use up the full 4 dB tolerance. It is commendable that the product meets its specification, but the specification isn't informative. What do the curve fluctuations indicate in the way of anything audible? We know that low-Q resonances are audible when they are revealed by humps of about 1 dB amplitude or more, so there are likely be some audible broadband spectral colorations. The smaller bumps might be evidence of acoustical interference, which is innocent, or resonances, which are not. Without more information we don't know which they are, but the suggestion is that this is not a neutral-sounding loudspeaker.

Figure 2.13(c), a spinorama, provides much more information. Resonances have been well controlled, because the bumps in the direct sound are not seen in the lower, spatially averaged curves. This is evidence of careful design. The specification applies only to the sound radiated on the reference axis, the on-axis direct sound from the loudspeaker. But we listen in reflective rooms, so off-axis performance also matters.

When one looks at the far off-axis performance, as displayed in the early reflections and sound power curves, there are large fluctuations that affect

the sound of the reflected sound field. The directivity index shows the rising directionality of the twin woofers, crossing over to the large 6-inch (150 mm) midrange driver around 400 Hz that, in turn, becomes increasingly directional until it crosses over to the unbaffled 1-inch (25 mm) tweeter around 4 kHz. The broad dip in sound power and the rise in directivity around 3 kHz are predictable results of the mismatch in directivity in the midrange-to-tweeter crossover region. Improved results could be obtained by using a shallow horn on the tweeter, so that it better matches the directivity of the midrange transducer at crossover, but that is not possible with the unbaffled tweeter in the chosen visual design. This lack of attention to off-axis performance is like that seen in Figure 2.6, another expensive British loudspeaker from 34 years earlier.

Figure 2.13(d) shows that the early-reflections curve from the spinorama is a good predictor of the steady-state spatially averaged room curve. If a room curve is the only data one has, it might be tempting to equalize it to a smooth, gently downward sloping shape—a popular target. However, doing so would make the important on-axis curve worse, and a smooth, flattish on-axis curve is the preferred target performance. The solution is a better loudspeaker. Room curves are results, not targets—except at bass frequencies.

This is all interesting technical talk, but how does it sound? Figure 2.14 shows the results of a subjective comparison between this and three other loudspeakers in the $8000 to $11,000 per pair price range. All four had been applauded by the audiophile press, with accolades like “editor’s choice”, “Class A”, “product of the year”, cover photographs, etc. Readers of these publications were led to believe that any one of these products would be a superb choice. The prices only enhance such a belief.

Figure 2.14 is another example of agreement between the subjective and objective domains. The results of double-blind listening tests indicate that two loudspeakers statistically tied for first place; one was a clear, statistically significant second; and one significantly brought up the rear. In looking at the corresponding measured data, it is not difficult to see the progressive decay in performance. There would seem to be no great

mystery here, but two key elements underlying this display are very rare in the audio business: properly conducted double-blind listening tests and accurate, comprehensive anechoic measurements.

> When expectations from published specifications differ significantly from audible reality, it is easy to understand how the old expression "we can't measure what we can hear" came to be. The problem is inadequate measurements, not the supremacy of the human hearing system.

Naturally, the subjective evaluations in the magazine reviews that praised these loudspeakers were conducted in the "take-it-home-and-listen-to-it", single-stimulus fashion, with the problems discussed in Section 1.2. Break-in was a theme in the *Stereophile* magazine review of loudspeaker M, which required several weeks before the anticipated sound qualities were fully revealed. The reviewer had an admitted "passion" for electrostatic loudspeakers, which may explain his patience (Greenhill and Atkinson, 2001). In contrast, the same person reviewed loudspeaker R and praised it enthusiastically. It sounded impressive from the outset, and break-in was not mentioned (Greenhill and Atkinson, 1999). What (or who) was "breaking in"? John Atkinson's measurements conveyed the same messages that those in Figure 2.14 do, even though they were limited by being approximations to full anechoic data. Useful measurements have been around for a long time, just not given adequate respect—even by some of the subjective reviewers sharing the pages.

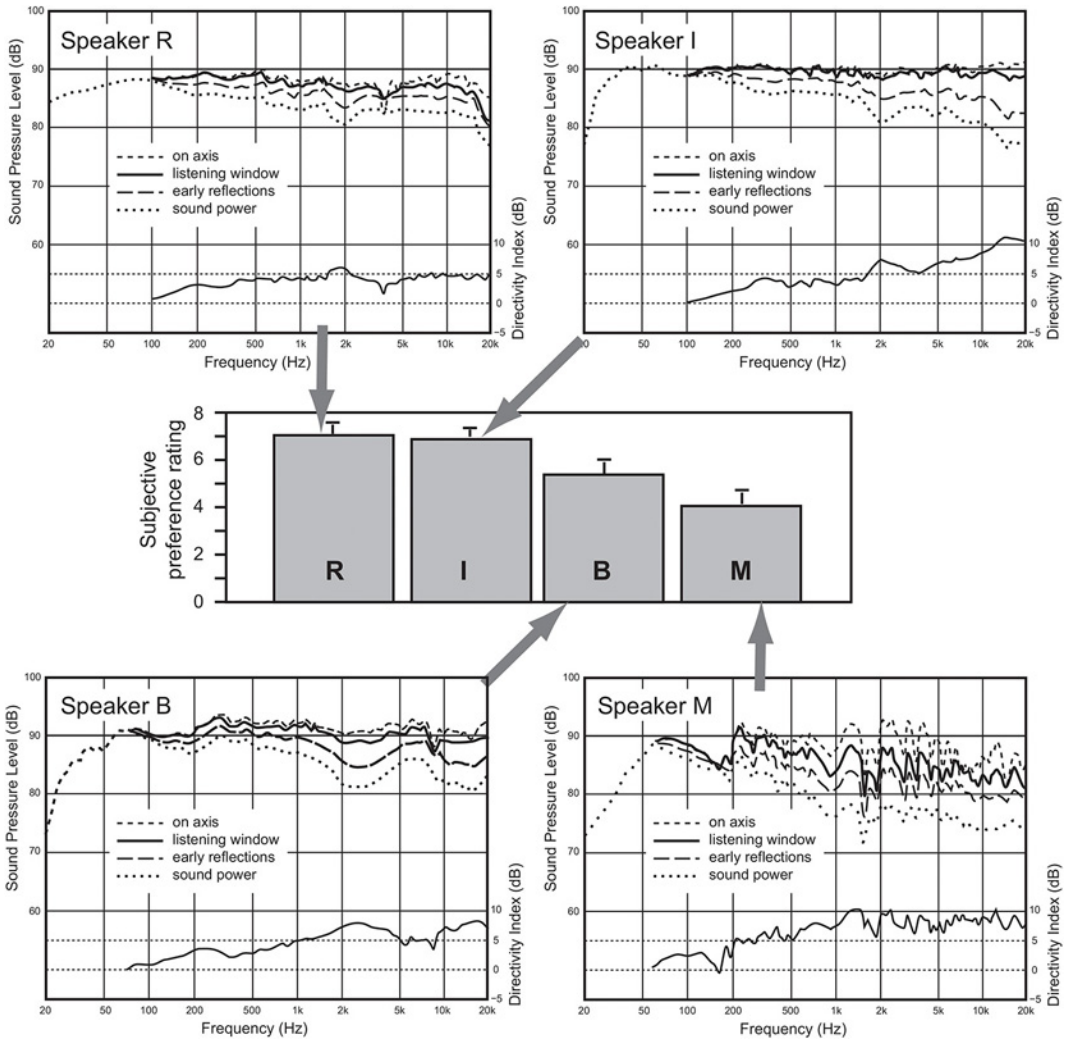


FIGURE 2.14 *Subjective ratings and spinoramas on four high-end loudspeakers. The error bars on the top of each histogram bar show 95% confidence levels, confirming that loudspeakers R and I are in a statistical tie for first place. The other differences are highly significant. Loudspeakers R, I and B are forward-firing cone-dome systems, but R also has a second rear-firing tweeter. The lowest-rated loudspeaker M is a hybrid system with an electrostatic mid-high frequency driver mated to a conventional woofer. Loudspeakers R, I and B are free from obvious resonances, and differ mainly in the constancy of their frequency responses and directivities. Loudspeaker M is very directional, and shows evidence of multiple resonances: bumps that appear in all the curves including the sound power. Loudspeaker R is a*

Revel Salon, I is an Infinity Prelude MTS, B is a B&W 802N, and M is a MartinLogan Prodigy. A pair of each model was tested for consistency. This is a modified version of figure 18.14 in the first edition of this book. ⏎

Recognizing true excellence or true inferiority in spinorama style measurements does not require any special training. Interpreting the relative merits of the vast middle ground of loudspeakers that don't do everything well is more difficult. Flaws come in many versions. The decisions come down to trading off problems and virtues, the audibility of which are likely to be dependent on program material, possibly the listening room and its arrangement, and perhaps the ability of individual listeners to hear or ignore certain kinds of defects. The subjective evaluations reported here were arrived at in an elaborate positional substitution listening room, where the contribution of the room to bass response was constant. This is not normally the case.

An interesting fact is that the B&W 802N is found in some high-profile music recording facilities and scoring stages. Some recording engineers believe it to be especially useful for classical music. It has been said that the sagging output around 3 kHz is compensation for excessive brightness in strings—the result of placing microphones above the violins where they project energetic high frequencies that are not heard in that proportion by live audiences (Meyer, 2009, figure 4.1). If this is the case, the remedy is to equalize the signal, not to use monitor loudspeakers with a compensating error, thereby contributing to the circle of confusion. In any event, this is a non-neutral loudspeaker being used in situations where neutrality should be a requirement.

In the past, everyone, audio professionals included, lacked meaningful technical guidance. Now there is no excuse. If manufacturers don't publish comprehensive anechoic data, ask for it. If they don't have it, or won't show it, take your business elsewhere.

With comprehensive and accurate anechoic measurement data, consumers who can interpret the curves are informed about the potential

sound quality of loudspeakers without hearing them. It is a skill worth acquiring as online purchasing is the new norm. As can already be seen, this book will help consumers to "read" the information in the spinoramas. It is an "analog" skill; neither mathematics nor an engineering degree is required. As stated at the beginning of this book, listening is the final arbiter of sound quality, but when adequately controlled listening situations are not available, measurements alone can provide guidance, perhaps avoiding expensive errors.

How loudspeakers will sound in the customer's home or recording control room has always involved an element of chance because bass performance is determined by the listening room and arrangements within it. Bass alone accounts for approximately 30% of our overall impression of sound quality, so clearly the job is not complete until something is done to alleviate the worst effects of room resonances. Chapter 14 focuses on this topic.

2.10 A CAREER DRAWS TO A CLOSE

As I reflect on my years of research work, none of it feels like a "job"; I was doing what interested me and I found myself in circumstances that enabled it to happen—and get paid for it. I had access throughout the decades to accurate, comprehensive anechoic measurements as well as controlled double-blind subjective evaluations of many loudspeakers using many listeners. These fulfilled the fundamental needs for productive psychoacoustic research, as described in Figure 1.1. Few people in the world have had this luxury. With these facilities at hand, psychoacoustic linkages and correlations appeared quickly and persuasively. Without them, workers have incomplete or unreliable data. The uncertainties and inconsistencies can lead to experimental errors, and opinions having more influence than they should.

A few minutes on the internet proves abundant evidence of strongly stated opinions and unimpressive, or missing data. Audio forums have seemingly endless debates about settled facts, but the facts are not in

evidence, or believed, so the debates go on. This is audio as described in the little graphic in the introduction, an awkward alliance of Art and Science, Opinions and Facts, two "cultures" that have communication problems. But for those who can mentally bridge the culture gap and have the patience to read, listen and learn, the audio scene becomes much more rational, manageable, predictable and most importantly, enjoyable.

I retired from Harman International in 2007, after which I was retained as a consultant until 2019—age 81. During this time, I continued informal contact with the engineers and researchers, delivered numerous tutorials in support of engineering and marketing, including a 75-minute overview of the core of the research delivered at McGill University CIRMMT, available on YouTube (Toole, 2015b). I wrote two versions of this book, the first and third editions. The second edition was a labeling error when a new publisher took control. I was widowed in 2019 and remarried in 2020. In 2024 my new wife and I left sunny southern California and moved back to our home of many years, Ottawa, Canada, to be closer to relatives, friends, audio industry colleagues from the formative NRCC research period, and to experience distinct seasons. It is where this book is being composed.

REFERENCES

ANSI/CTA-2034-A R-2020 (2015). "Standard Method of Measurement for In-Home Loudspeakers", *Consumer Technology Association, Technology and Standards Dept*. www.CTA.tech. ⏎

Devantier, A. (2002). "Characterizing the Amplitude Response of Loudspeaker Systems", *113th Convention, Audio Eng. Soc*., Preprint 5638. ⏎

Greenhill, L. and Atkinson, J. (1999). "Revel Salon Loudspeaker", *Stereophile*, May 28. www.stereophile.com/floorloudspeakers/96/index.html ⏎

Greenhill, L. and Atkinson, J. (2001). "MartinLogan Prodigy Loudspeaker", *Stereophile*, July 12. www.stereophile.com/floorloudspeakers/390/index.html ⏎

IEC 60268-13 (1998). Technical Report—Type 3, "Sound System Equipment—Part 13: Listening Tests on Loudspeakers", *International Electrotechnical Commission*, Geneva, Switzerland. ⏎

Meyer, Jürgen (2009). *Acoustics and the Performance of Music*, 5th edition, Springer. ⏎

Olive, S. E. (2004a). "A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 1—Listening Test Results", *116th Convention, Audio Eng. Soc.*, Preprint 6113. ⏎

Olive, S. E. (2004b). "A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 2—Development of the Model", *117th Convention, Audio Eng. Soc.*, Preprint 6190. ⏎

Toole, F. E. (1982). "Listening Tests—Turning Opinion into Fact", *J. Audio Eng. Soc.*, vol. 30, pp. 431–445. ⏎

Toole, F. E. (1985). "Subjective Measurements of Loudspeaker Sound Quality and Listener Preferences", *J. Audio Eng. Soc.*, vol. 33. pp. 2–31. ⏎

Toole, F. E. (1986). "Loudspeaker Measurements and Their Relationship to Listener Preferences", *J. Audio Eng. Soc.*, vol. 34, pt. 1, pp. 227–235, pt. 2, pp. 323–348. ⏎

Toole, F. E. (2006). "Loudspeakers and Rooms for Sound Reproduction—A Scientific Review", *J. Audio Eng. Soc.*, vol. 54, pp. 451−476. ⏎

Toole, F. E. (2015a). "The Measurement and Calibration of Sound Reproducing Systems", *J. Audio Eng. Soc.*, vol. 63, pp. 512–541. This is an open-access paper available to non-members at www.aes.org www.aes.org/e-lib/browse.cfm?elib=17839 ⏎

Toole, F. E. (2015b). "Art and Science/Opinions and Facts", *CIRMMT Distinguished Lectures in the Science and Technology of Music*, McGill University, Montreal. www.youtube.com/watch?v=zrpUDuUtxPM&t=339s ⏎

CHAPTE
R 3

Perspectives on Sound Production, Recording and Reproduction

Floyd Toole

DOI: 10.4324/9781003477495-4

The origins of modern acoustic science lie in the domain of large auditoriums, including concert halls, beginning with the concept of reverberation time by Wallace Clement Sabine at the end of the 19th century. Whether classical music is appealing or not, the reverberation, spaciousness and envelopment generated in live concert hall performances have widespread appeal. Recording engineers can capture components of real spatial cues, or they have electronic synthesizers allowing these dimensions to be incorporated into any kind of music and movies, adding to the artistic palette. These sounds are mixed and manipulated in many ways, and finally approved after listening in recording control and mastering rooms. The “live” performances that matter for recordings occur in small rooms equipped with loudspeakers—sound reproduction. It is reasonable to

think that the professional side of the audio industry can provide guidance to consumers as to how best to enjoy the results of their efforts. Professional recording engineers are concerned about how their recordings "translate" into the real world, which is variable. But their own world is also variable. The topic of how to improve the odds of satisfactory "translation" of the art was introduced in Figure 1.8 and will appear repeatedly in this book.

3.1 SOUND PRODUCTION

A live concert performance is special. Upon walking into the hall, I am aware of a large space—I hear the space that my eyes see. This does not happen when listening to classical recordings at home. As the musicians tune up and practice difficult passages the timbres are enriched by countless reflections—repetitions that give the hearing system more opportunities to hear timbral subtleties. I can localize individual musicians, but even though the sound begins at those locations the timbres linger in the decaying reverberation all around me. When the music starts, it all magnificently comes together. The sound of the hall is an inseparable part of the performance: rich timbres combined with enveloping space. I am "in" the performance. This is a complex listening experience. In some ways it is a "reference", but it is important to note that with different musicians, instruments, conductors and halls, the "reference" is really not a constant entity.

Concert halls, those palaces of acoustical culture, are all different, such is the vanity of architects and acoustical designers, influenced by a need for the facility to be profitable. Each one is a very expensive experiment, and those that succeed are wonderful. Nevertheless, the essential identifying timbres of voices and musical instruments remain remarkably constant. We have considerable ability to separate the sound of the source from the sound of the hall. In other words, we appear to adapt to the room we are in and "listen through" it to hear the sound sources. Bregman (1999) calls this "auditory scene analysis". We seem to be able to perceptually "stream" the sound of the voices and instruments as significantly separate from each

other and from the sound of the room. It parallels the concept shown in Figure 1.6.

As pleasurable as the classics are, the majority of our entertainment falls into the numerous subdivisions of what is collectively called popular music. The recording methods used are also shared with most jazz we hear. In general, what we listen to is captured by microphones located very close to voices and musical instruments, or electrically captured without any acoustical connection, and then mixed and manipulated in the control room of a recording studio. The "performance", the art, is what is heard through the monitor loudspeakers during the mixing and mastering operations.

Big name artists engage in elaborate tours, spanning the globe. They became popular through recordings created in studios and reconstructing the essence of that recorded "sound" in a concert situation is what the fans expect. Portions of studio recordings are frequently blended into live performances, and lip syncing is apparently not uncommon. It might seem like cheating, but there are some effects that cannot be duplicated in live performances. In the end, the delivered "art" is what matters.

Figure 3.1 shows that in live, amplified/sound reinforced, popular music performances, the front-of-house (FOH) engineer is in control of the performance heard by the audience. There is a parallel audio system in the form of front-of-stage monitor loudspeakers and/or in-ear monitors worn by performers, all intended to help the musicians hear each other and any prerecorded components. It is an artistic creation in real time and can vary enormously with different engineers, their tastes and, interestingly enough, how well they hear. I left one concert at intermission because the sound was far too loud, and not good sounding, yet the loudspeakers were excellent products. I learned later that the FOH engineer was known by insiders to have serious hearing loss but had a long relationship with the performer. What a waste.

3.2 SOUND REPRODUCTION: THE AUDIO INDUSTRY

Most of our listening experiences involve music recordings, movies, broadcasts or streaming audio that are reproduced through loudspeakers in cinemas, listening rooms, cars, or through headphones. This is the audio industry, as shown in Figure 3.2.

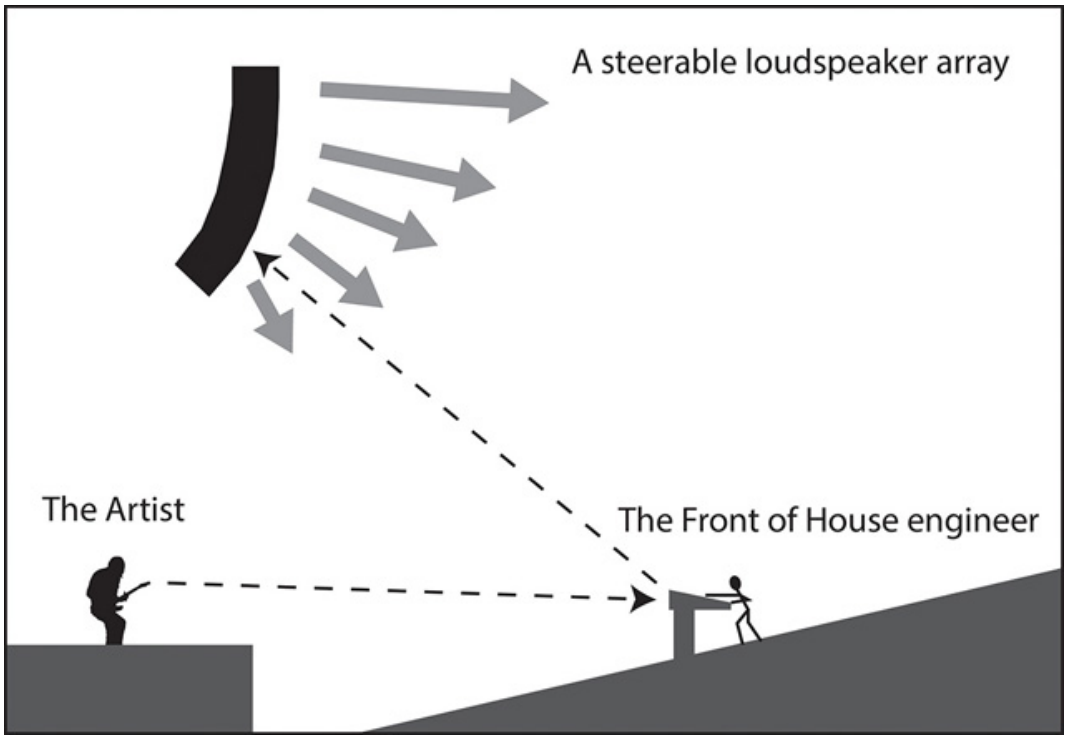


FIGURE 3.1 *A functional diagram of a tour sound system. The microphone and direct-wired inputs from the artists on stage are mixed and manipulated by the front-of-house (FOH) engineer based on what is heard from the loudspeaker array, the sound of which is shared with the audience.*

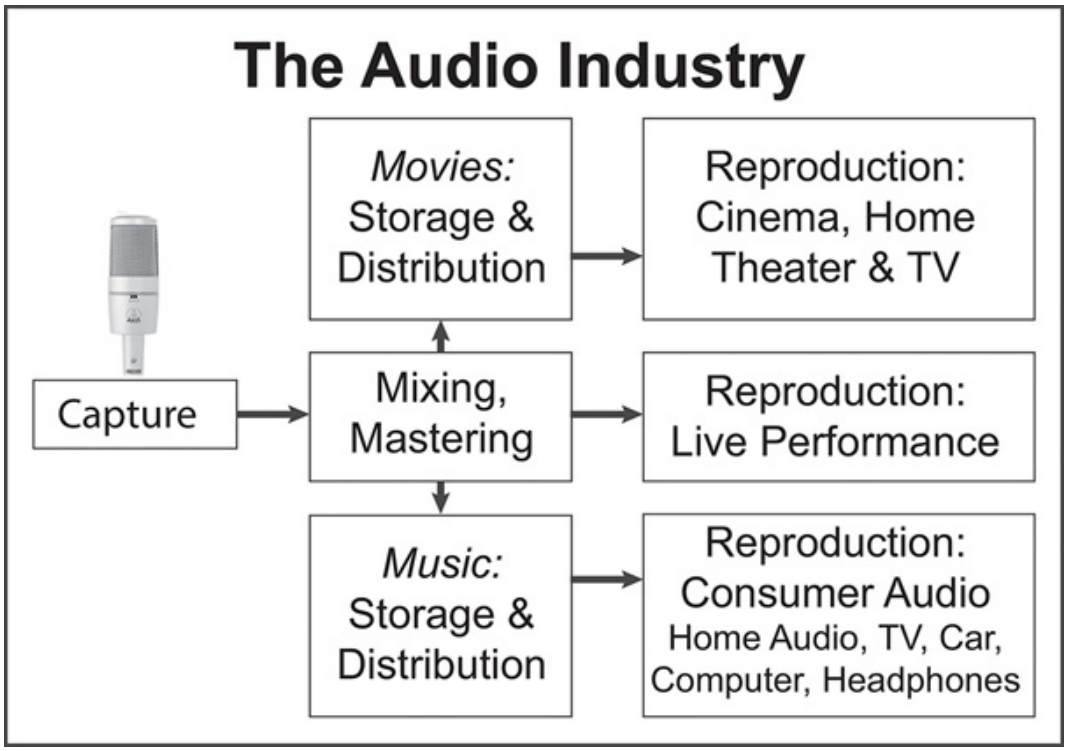


FIGURE 3.2 *The audio industry as we know it.*

The audio industry is a complex operation, requiring extensive standardization if it is to work. All the devices within these different activities and business units need to be electrically compatible with the signals moving through them. Equally important is the matter of what listeners in these varied circumstances hear.

Do the recordings "translate" among facilities within the professional system, and then to the consumers? Unfortunately, not often enough. The required standardization is not in place. This is an industry populated with creative artists, and not all of them appreciate the need for technical standardization of equipment before using it to create their free-form artistry. One reason seems to be a lack of trust in measurements of loudspeakers and rooms—the artist and the scientist not communicating.

All too often, monitor loudspeakers are chosen because they color the sound in a way the recording engineer likes. Their loudspeakers become part of the equalization applied to the recording, but only they will ever hear the art as it was created; translation is already compromised. If a

particular coloration is desired in a recording, use an equalizer, so those of us with neutral loudspeakers and headphones will hear it. It sounds simple. But really they have been victims of circumstance, because, as will be shown, monitor loudspeakers have a long history of being powerful and unbreakable, but also often very colored. Neutrality was not an option. That has changed.

3.2.1 Microphones—The First Transducers

This book is focused on the last transducers, loudspeakers that convert electrical energy into sound. But the entire audio industry begins with microphones that convert sound energy into electrical signals. Variations in microphone performance are real, expected and part of the artistic creative process. Whatever the intent of the recording engineers, it is helpful to have trustworthy measurement data, in a consistent form for comparison purposes, to assist in making choices.

There are reputable texts on microphones that explain the fundamentals of microphones and how they are used (e.g. Rayburn, 2011, an update of the classic John Eargle text). Structurally, a microphone is simpler than a loudspeaker, many having only a single diaphragm that is smaller than most tweeters, although very directional microphones can be quite complex. Most are free from the bothersome resonances that can exist in loudspeaker systems. Even so, Sean Olive and I thought that some consistently recognizable audible problems at high frequencies might be attributable to the recording microphones, not the tweeters in loudspeakers being evaluated.

Several popular recording microphones were measured, and several showed significant variations in high-frequency performance, bass response and irregularities in directivity, all of which color the direct and reflected sounds that are recorded (Olive and Toole, 1989a). The measurements paralleled the technique used for describing loudspeaker performance, covering 360° orbits and allowing directivity index calculations. As the authors noted, there are microphones that are more “artistic” than “accurate”, which is fine so long as the users of the microphones have prior

knowledge of their performance and can make rational decisions about their use. Unfortunately, the microphone industry is inconsistent in how it describes the performance of its products, making comparisons difficult.

The data shown in the paper reveal performances of all microphones in the same format using a measurement method that parallels that used for loudspeakers and used the same scaling of the data. It is interesting to note that the variations in microphones are of the same magnitude as those seen among loudspeakers. Some microphones showed worse high-frequency performance than tweeters in good loudspeakers. There are large differences among widely used "pedigreed" products. Naturally, microphone performance is subjectively evaluated by listening through loudspeakers in a room, so they are a component in the "circles of confusion" shown in Figures 1.4 and 1.5.

Subjective evaluations of loudspeakers and other audio equipment therefore inevitably incorporates a measure of microphone evaluation, yet this is a factor frequently not recognized. A second part of this paper intended to address this aspect, but other work took priority, and it was never completed. Preliminary tests indicated that the choice of microphone in a recording could affect subjective ratings of loudspeakers in predictable ways, an unsurprising result, as they are part of the circle of confusion—unless its peculiarities are compensated for by the recording engineer.

It is reassuring to know that, over most of the frequency range, many modern microphones used in recording studios these days are fundamentally resonance-free neutral transducers. Recording engineers often have collections of microphones, some historic, having different timbral signatures, proximity effects, dynamic ranges, etc. that are used in specific applications. They are part of the art. The microphones used in live performances are generally less than neutral but offer the compensation of being able to survive being dropped and to sound respectable with a vocalist almost chewing on the windscreen. The FOH engineer is responsible for making the artist sound good.

A major problem with microphone pickup of sound is that sound is captured at a point in space, and instruments and voices radiate different

sounds in different directions. In a live *acoustic* concert, all these sounds eventually reach our ears through reflections in the performance space, but a microphone positioned in an audience location collects sound that is judged to be too reverberant. Two ears and a brain—binaural hearing—in the same location can perceptually separate the direct and reflected sounds, but the microphone simply combines them. Therefore, recording microphones are placed much closer to the voices and instruments, meaning that only a sample of the sound from the source is recorded, which may or may not be a good representation of the sound heard in the audience. Compromise is necessary, and engineers have worked out schemes that work for them, with or without adding equalization. Recording engineers can manipulate the mix to sound something like the live event, to create a totally artificial experience, or anything in between. In the absence of live references, contemporary listeners have adapted to the new close-miked instrumental sounds which may have more clarity and detail than a more distant "acoustical" version.

3.2.2 Reproduction Venues

The objective of this book is to understand the relationship between loudspeakers, rooms, headphones, and listeners. It is essential to first understand what is expected of these devices as they are used in the audio industry and in homes. Does one size fit all?

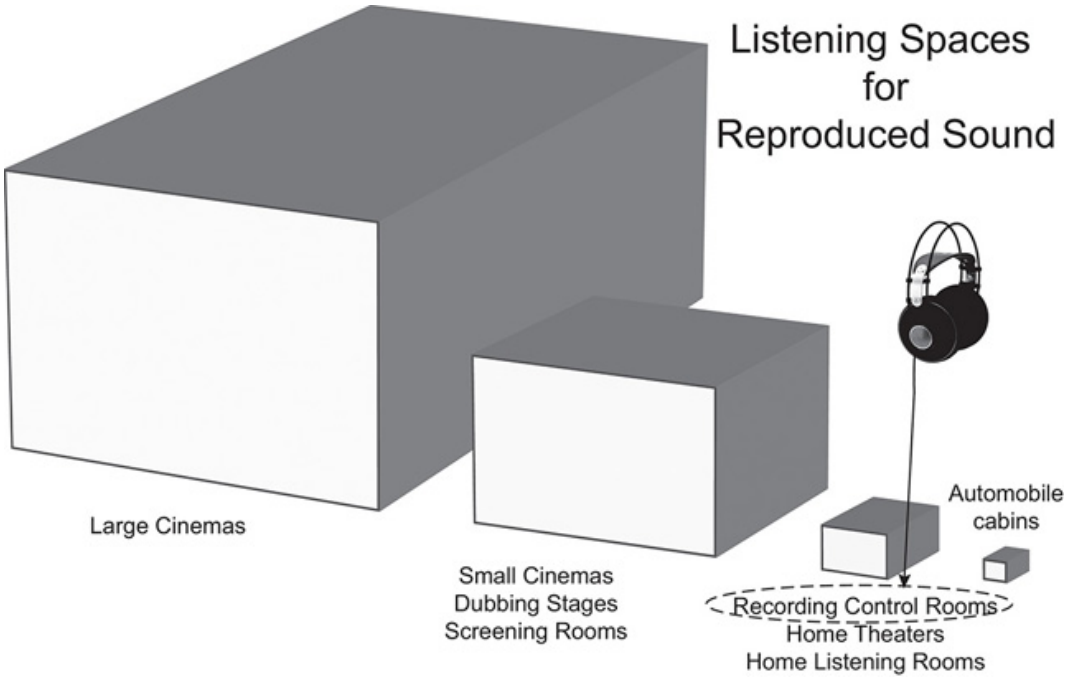


FIGURE 3.3 *The range of listening spaces within which accurate sound reproduction is desired.* ⏎

Looking at Figure 3.3, the answer is obviously "no".

The reality is not as bad as the illustration suggests because cinemas and dubbing stages have reverberation times that are not very different from domestic rooms. The directivity-controlled loudspeakers combined with generous acoustical treatment in cinemas mean that listeners end up in basically similar direct and reflected sound fields. However, the loudspeakers required to deliver the required sound power into large spaces are very different devices from those in smaller rooms. Although recording and mastering rooms, home theaters and domestic stereo listening rooms share the same size category, they cover a wide range of internal acoustical behavior.

Headphones normally replay stereo soundtracks created using loudspeakers in recording control rooms. Although it is possible to achieve a good timbral match, the spatial presentation is fundamentally different from that delivered by loudspeakers.

The bar-graph portrayal in Figure 3.4 could very well be wrong in detail, but there is little doubt that the trend is correct. The bulk of scientific research has been devoted to performance spaces, concert halls. A fraction of that attention has been paid to workspaces, homes, cars and headphones. However, the time spent listening in these modes shows almost the opposite trend. The circumstances within which we are primarily entertained by audio programs have not had the proportional benefit of scientific research. As a result, opinions, traditions, fashions and folklore have had and still have significant influence in these domains.

Let us begin this discussion by examining some possible objectives. What sound do we want to reproduce?

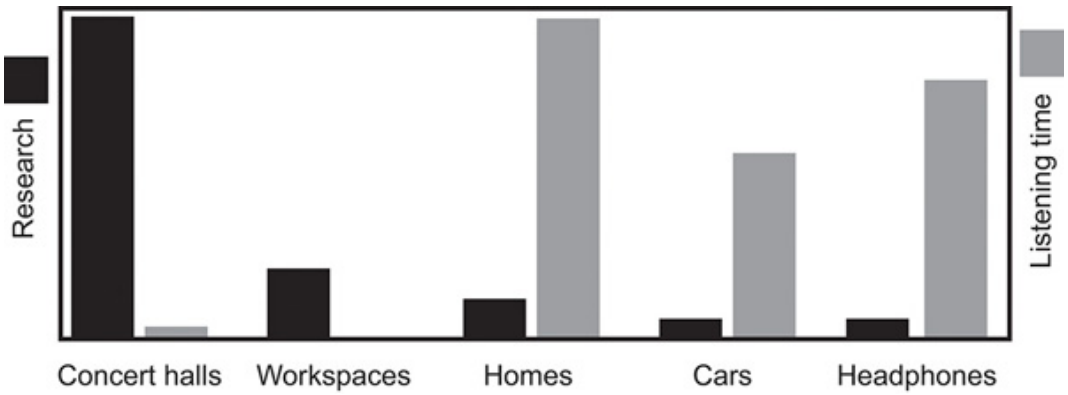


FIGURE 3.4 *A crude estimate of the quantity of scientific research devoted to various acoustical spaces and listening situations compared to the amount of time spent listening in each of them.*

3.3 CAN WE REPRODUCE "REALITY"? BINAURAL TECHNIQUES

The most ambitious objective would be to capture and reproduce "reality"—persuading listeners that they are in a live performance at a different time in a different place, immersed in the acoustical ambiance of the venue. Morrison (2009) gives a good historical review from a subjective audio perspective. It has been a longstanding goal among audio enthusiasts.

The following, from a technical point of view, is an analysis of the factors involved.

There are two approaches to achieving this:

1. Capturing and storing the details of the three-dimensional sound field of a live performance and reconstructing it in the customer's listening space.
2. Capturing, storing and reproducing the sounds that would have arrived at the customer's ears if they had been at the original performance.

Option 1 is probably impossible. However, if it could be done, it would entertain multiple listeners, allowing movement within the reconstructed sound field. In the 1930s, Bell Labs of AT&T conducted an ambitious experiment for the day when they transmitted a "performance" from Philadelphia to Washington, DC, by setting up a linear array of microphones across the front of the stage in the performance venue and transmitting their signals to loudspeakers set up across the stage front of the reconstruction venue. This was an attempt to transmit the "soundstage" only, letting the natural acoustics of the receiving venue provide the rest of the acoustical experience. They found that many channels were most persuasive, but that three (left, center and right) were a good approximation. Further, they thought that binaural audio would be a good alternative (Steinberg and Snow, 1934).

The present-day multichannel "immersive" systems are fundamentally different. They don't attempt to capture and reconstruct a sound field, but they can deliver clues suggesting the basic perceptions of a live experience. This will be discussed later.

Option 2: The pioneers of the 1930s were prescient. Binaural recording and reproduction—head-related audio—works very well, but only for a single listener. This may seem like a serious restriction, but stereo too is designed for a single listener in the sweet spot. Capturing the complex patterns of direct and reflected sounds as they arrive at the listening location

is essential because our binaural hearing system is sensitive to the direction and timing of incoming sounds. To deliver a perceptual duplicate of a live performance, the ears must receive the same sounds that would have been heard in a seat at the live event. Each direction of incoming sound is characterized by an interaural time and amplitude difference, modified in a distinctive way by the detailed shapes of the head, the external ears and the torso. These modifications are known as HRTFs, or head related transfer functions.

The microphone required to capture these directional and spatial details is an anatomically correct dummy head and torso, a mannequin. Tiny microphones are placed at the eardrum locations at the ends of ear canals inside the dummy head, including an acoustical termination that simulates the sound reflecting properties of a real eardrum. Because we are all anatomically different, the external ears (pinnae) are interchangeable and can even be customized to match a particular listener. These head and torso simulators (HATS) are widely used for measuring headphones (Chapter 15) and for various other research purposes. Professional audio versions, such as the Neumann KU100, exist for creating binaural recordings, and they tend to be simplified in various ways, like placing the microphone at the entrance to the ear canal, and not including a torso. Very small microphones present challenges with respect to background noise, a key factor in commercial recordings.

For obvious reasons, the ideal playback of binaural recordings is through accurate headphones, so that the recorded sounds are delivered to the correct eardrums with minimal modification. Insert headphones deliver sound directly into the ear canal, avoiding complications of personal differences in the fitting of on, over or around-the-ear styles of headphones to the external auditory apparatus. Bass variations due to leakage, and external ear influences are persistent problems (Toole, 1976, 1984). There will be much more on this topic in Chapter 15.

The results can be impressive, especially if *head-position tracking* is implemented so that the band does not move with the head, allowing the perceived sounds to be persuasively external. There is an issue of the

detailed shapes of ears and torso of the dummy being different from those of any individual listener—our bodies are all distinctive—so it is not surprising to hear reports of more realistic externalization of the sound when personalized HRTFs are employed (e.g. Griesinger, 2019).

3.3.1 Head Related Transfer Functions: HRTFs

The complex shapes of the external ears, the pinnae, combined with shapes of the head and torso modify the sounds arriving at the ear canals differently for sounds arriving from different directions. This spatial encoding can be measured with tiny microphones in the ear canal or at the eardrum so that we know how the sounds originating in different directions are modified on their way to the ears.

The measurements are ideally made in an anechoic space using loudspeakers at specific vertical and horizontal angles with respect to the head. The data are in the form of the complex transfer function—the amplitude and phase vs. frequency responses, or the equivalent transformed version, the impulse response for each angle for incoming sound—hence the name.

Significant problems are that the default localization in headphone listening is inside the head or behind, and that rotating the head takes the recorded sound image, the band, with it. Clearly the brain is displeased with the information it is receiving. HRTF information that is inappropriate for individual listeners is often blamed, as is the absence of head tracking that alters the binaural signals as the head rotates, leaving the image locations unchanged. Figure 3.5 shows a solution.

The default perception might have its origins in a human defense mechanism in which sounds that are not reliably identified or are from an invisible source may be perceived to be inside the head or behind—anywhere outside the field of view. The head spontaneously moves, and the correct direction is revealed but only if the sound continues. An isolated twig “snap” no doubt presented anxious moments for our hunter-gatherer ancestors.

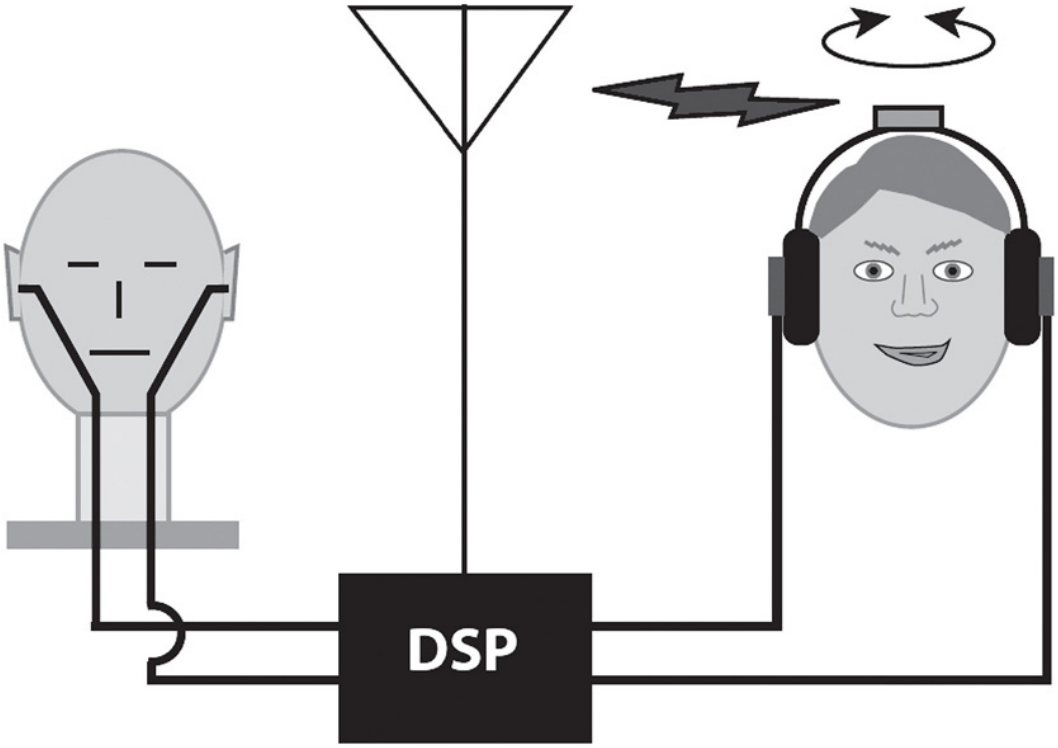


FIGURE 3.5 *A binaural system in which head position is interpreted by a sensor and appropriate real-time changes are made to the binaural signals. The listener experiences a stable soundstage, the musicians don't move, even while turning the head. From Toole (1998).* ⏎

Head movements have long been assumed to be important to satisfactory externalization in binaural headphone reproduction (Toole, 1970). The brain protests when the musicians and the performance venue move with the head. But mismatched HRTFs are a logical contributing factor to the perceptual error.

For headphone listening, signals can be processed using binaural HRTFs to create the impression of sounds arriving from different directions—constructing a three-dimensional illusion. These synthesized localizations are widely used in gaming and in processing stereo and multichannel signals for binaural headphone presentation. See Figure 3.6.

3.3.2 Anecdotal Observations of Adaptation at Work

HRTFs are different for every individual, which means that as individuals we have adapted to—normalized—our personal HRTFs. We all are able to recognize the timbres of familiar sounds, even though the same sounds do not reach all of our eardrums. In sound reproduction we recognize the uncolored sound from loudspeakers having flat, neutral frequency responses, while listening through our hearing system HRTFs that are anything but flat. We start auditory adaptation from birth, it seems. An interesting question is: how flexible is that adaptation?

From a strict hard science point of view, improving the match between a listener's HRTFs and those involved in capturing the recorded event should make everything better—sound quality and external localizations should improve. Compensating for head movements will still likely be a requirement. This is an area of continuing research. However logical this approach sounds, there are experiences that challenge that simplicity, indicating that there is more to be considered, adaptation in particular.

In everyday life we often listen through modified HRTFs. Wearing a sun hat or winter knitted hats that cover the ears, parka hoods, earmuffs and other paraphernalia, even deliberately distorting the fleshy pinnae, all modify HRTFs. Sitting with our head close to or on a chair headrest makes major changes to the sounds arriving at our ears, but we normally don't notice audible distortions. Sound quality and localization continue in a relatively normal fashion. Humans seem to adapt in real time. and part of the reason may be that a combination of visible sound sources and head movements enable revised correlation between the modified HRTFs and real sound source locations. We might be adapting, in real time. Plausibility is a likely factor; that is, we tend to ignore cues that are implausible or contradictory.

And then there is the effect of the listening space. How readily does the brain accept the sound of one venue while listening in another? The visual cues do not match.

In the 1970s Sennheiser had a simple mass-market binaural microphone product, positioning small microphone capsules just outside the ears, but not in them. The encoded HRTFs had large errors. The headphones in use

were of questionable performance, so the HRTFs were further degraded. Recordings made in one venue and played back in a different one were entertaining but not convincing. In-head localization was the norm. However, out of interest, I sat in my stereo seat wearing the binaural microphone and made a recording of my stereo system playing. I then put on headphones, sat in the same stereo seat and played the binaural recording. It was uncanny. The beginning of the tape presented the household background noises as they were before putting the headphones on. It was as though nothing had changed. When the music began it was convincingly externalized—remarkable. The acoustical and visual contexts matched. Rotating the head destroyed the effect. An intriguing thought: playing a binaural recording of a classical concert while sitting in the same seat after everyone has gone would likely be very convincing, but only for the very privileged.

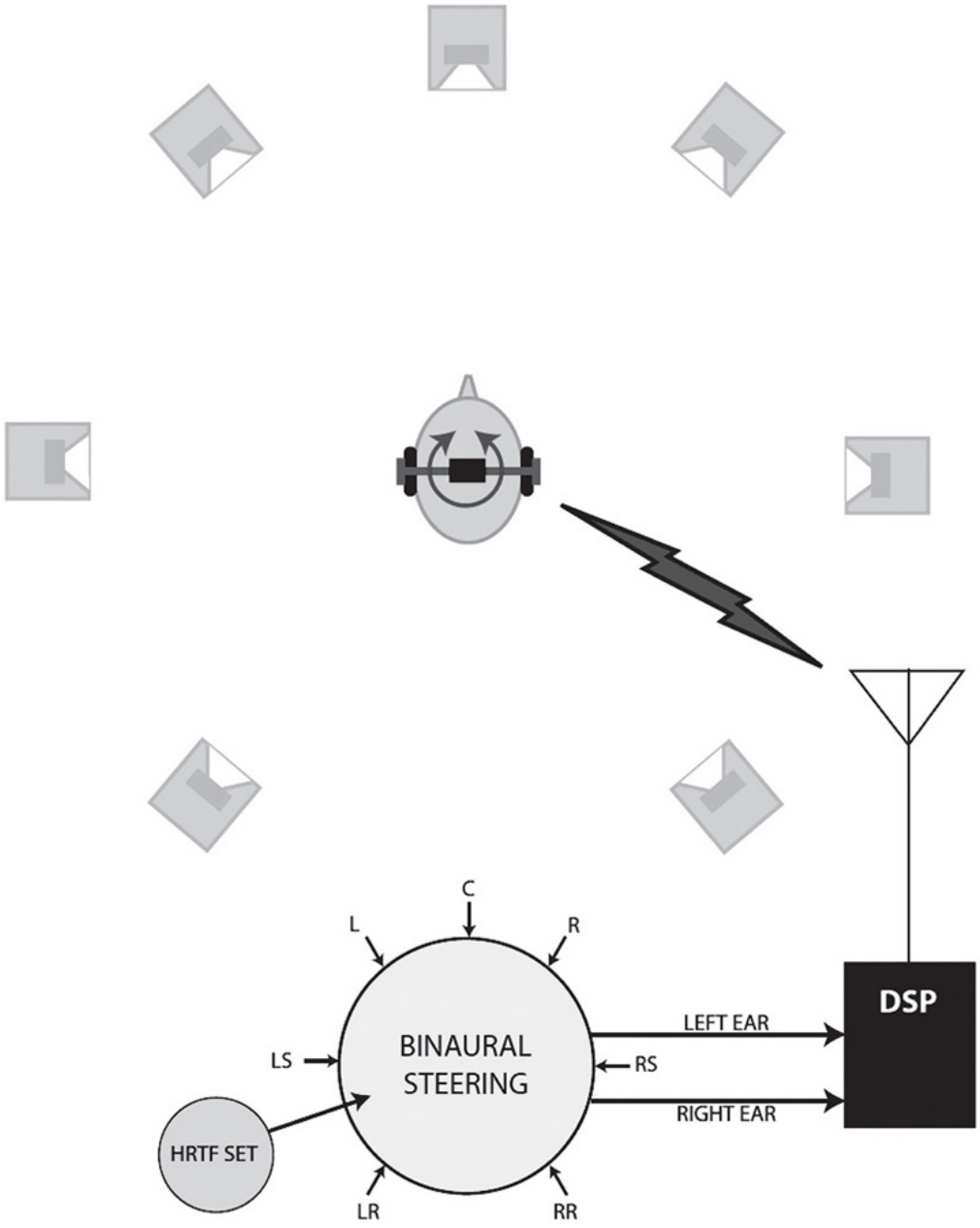


FIGURE 3.6 *A headphone home theater simulation. Head position tracking is combined with binaural digital processing so that the perceived directions of sounds remain fixed in space as the head rotates. Frontal sounds are correctly externalized, at a distance. In-head localizations and front/back reversals are eliminated. A binaural steering algorithm creates the left and right ear signals corresponding to the selected angular locations*

of some number of surround channels. Here seven are shown, but more are possible, including elevation channels. From Toole (1998). ⏎

A few years later I visited a researcher with common interests: Prof. Jens Blauert in Bochum, Germany. He and his students had just done some experiments in which successful image externalization was reported, attributed to improved matching of personalized HRTFs, but noting that several listeners shared the experience with the same HRTFs. He demonstrated it for me, and, to his dismay, it did not work for either of us. His conclusion was that the setup had been relocated to a corner of a lab; it was not in the original record/reproduction location. The acoustical and visual contexts did not match, and his adaptation had faded. For me it was all unfamiliar, so the experience ended up inside my head.

One of the most unexpected experiences occurred in the NRCC lab, where ear defenders had been developed and there was a continuing interest. Acos, a British company known for its crystal microphones, had created a clever device called the Erdefender (in the 1970s, as I recall). It consisted of a hard-shell ear defender for each ear with 1-inch crystal microphones on the outside of each aimed at 90° left and right to the sides. Under each shell was a small loudspeaker, and in the headband was an amplifier that transmitted the microphone signals to the respective ears. It was intended to be worn by workers in noisy environments, and it would allow normal communication with the outside world but limited the maximum sound level inside the hard shell to protect hearing.

Considering what was known about binaural hearing, it was a very crude device. Dr. Edgar Shaw was involved with the evaluation, and he was then in the process of explaining the physical acoustics of the external ears—pioneering research predating HRTFs (e.g. Shaw, 1966, 1974). Our expectations were not high for any "fidelity" in the experience, focusing on the hearing protection aspect. The sound quality was poor, the result of inexpensive directional microphones, questionable loudspeakers, and unknown signal processing. But we were surprised to find that after only a

few minutes of adaptation we all experienced externalization of the perceived images.

I wore the device for a working day and experienced no difficulty or disorientation other than occasional front-back reversals. These occurred when the head was stable—sitting at my desk, for example—and could easily be produced in the anechoic chamber. Head movements and compatible aural and visual information were of course responsible for enabling the adaptation. There were only rudimentary interaural amplitude and time cues and no HRTFs in the signal path. One's "ears", the simple microphones, were about 4 inches farther apart than the real ears attached to the top of hard-shell ear defenders. The conclusion was that if the acoustical and visual contexts match, and if head movements generate approximations of the expected interaural changes in sound at the ears, human adaptation can do a remarkable job of bringing the perceived binaural experience close to normal. With modern miniaturization and signal processing the concept of the Erdefender continues in the transparency modes of some insert headphones where elements of HRTFs are present but distorted.

3.3.3 Adding Credible Visuals Makes a Difference

In contrast to the last example in the previous section, listening to binaural audio without supporting visuals and head movement correlations is not a credible experience. Sounds are perceived to be inside, behind or close to the head. Well-synchronized head tracking, which keeps the sound images stable as the head moves, is a major improvement to externalization, but the visual information is incompatible.

Virtual reality (VR) headsets deliver sounds directly to the ears using headphones or near-field transducers—binaural audio—and the high-resolution displays covering much of the field of view can provide visual confirmation of the setting. If done well, a VR presentation, with synchronized binaural audio and head tracking, should be able to deliver a reasonable sense of "being there" for almost any kind of live event. It is possible that this has been done, but I have not experienced it.

Woodcock et al. (2019) investigated the interaction of audio and video components in 2-, 5- and 14-channel loudspeaker systems. They found that the strongest interactions were in the categories realism, sense of space and spatial clarity, and more effective as the channel count increased. The effect on stereo reproduction was negligible. This study used a 42.5-inch video display. This and smaller screens used in earlier experiments limits what can be learned. I have lived with a 10-foot (3 m) projection screen, and it alone creates a sense of "being" somewhere.

3.3.4 Binaural Crosstalk-Cancelled Loudspeaker Presentations

Loudspeaker-reproduced sound is always externalized but normally cannot deliver the binaural information separately to the ears. However, it is possible to process the left and right signals radiated by the loudspeakers so that the direct sounds arriving at the ears of a listener in the sweet spot are very close approximations to correct binaural signals—the acoustical crosstalk has been cancelled. When this is done well, the left stereo loudspeaker delivers sound almost exclusively to the left ear and the right stereo loudspeaker to the right ear. See Figure 3.7.

In this system a listener seated in the sweet spot is treated to a facsimile of binaural headphone listening, with an important difference: all sounds are perceived as external, because they are generated externally from visible sources. Normal small head movements are tolerated, and the listener can experience convincing sound localizations in directions outside the normal stereo soundstage, up to ±90° being very persuasive. Most listeners are no longer aware of the loudspeakers themselves, they "disappear", and the sense of distance is dominated by cues in the recordings. Obviously, the crosstalk cancelling process involves the HRTFs and, lacking such information for individual listeners, assumptions are made. The success of the cancellation is uncertain. Finally, room reflected sounds corrupt the illusion, so it is necessary to create a listening situation in which the direct sound is dominant. The effect is very persuasive in "near-field" listening situations, as at a computer.

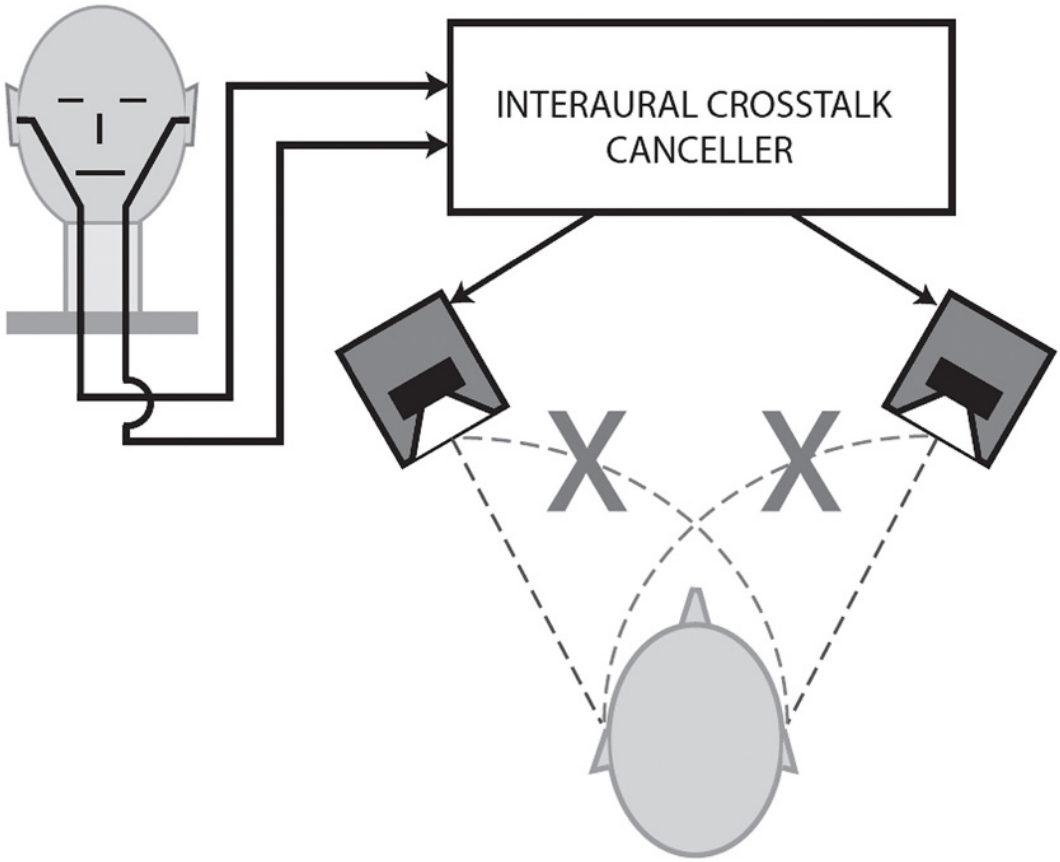


FIGURE 3.7 *A binaural system in which the signals are electronically processed in a manner that results in a cancellation of the acoustical crosstalk when all of the sounds arrive at the ears. The left loudspeaker then communicates only to the left ear and the right loudspeaker communicates only to the right ear. The crosstalk cancellation is not perfect, but it can be extremely effective. Obviously, the cancellation processing only works when the listener is exactly in the sweet spot, and different processing is necessary for different loudspeaker angular arrangements relative to the listener. From Toole (1998).* ⏎

Although intended for the reproduction of binaural, head-related recordings, listening to conventional stereo recordings is possible, in which case, what does one hear? It depends on how successful the crosstalk cancellation is, because what is happening is that hard-panned L and R images are delivered to the respective ears as isolated signals. The

perception can be of images to the sides of the listener. In the extreme, the normal ±30° stereo soundstage expands up to ±90°, which is not subtle. For large classical works it might be considered exaggerated. Popular music can sometimes generate isolated musicians well away from where one normally expects the band to be. It is a sound effect, and personal taste prevails.

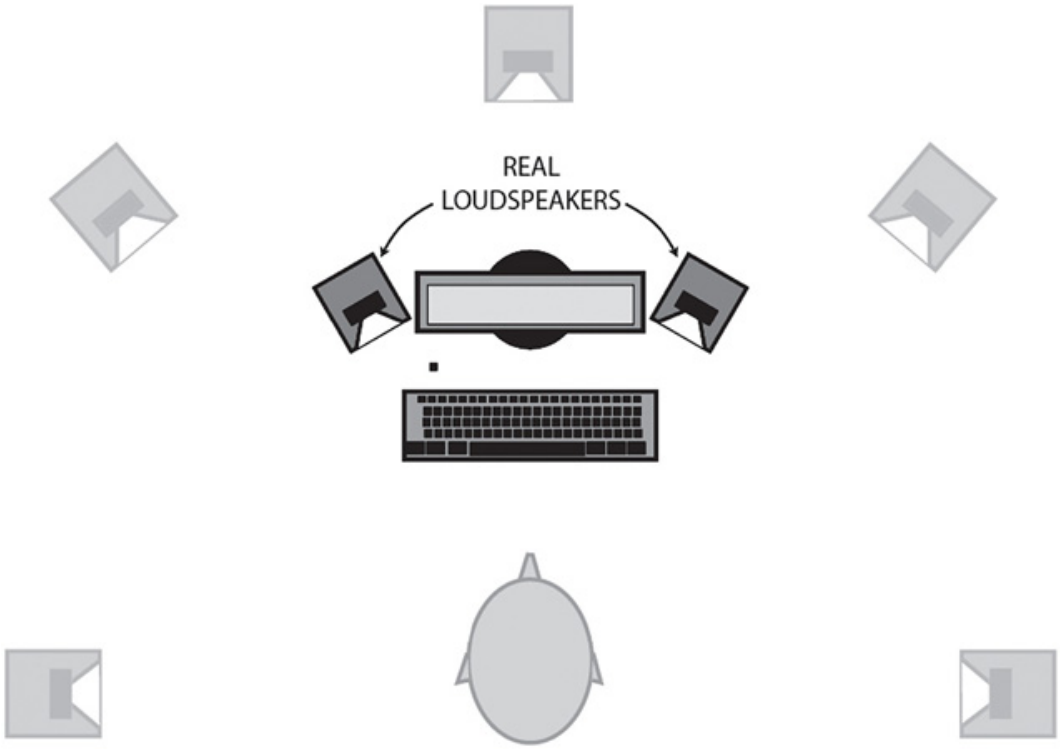


FIGURE 3.8 *Binaurally steered phantom loudspeakers can be used to reproduce the five channels in a multichannel format. From Toole (1998).*

Over the years there have been various versions of crosstalk cancellers. The "Sonic Hologram" (Carver, 1982), was a simplified version of binaural crosstalk cancellation in the electronic signal path, while the Polk SDA (Stereo Dimensional Array) loudspeakers tried it at the sound reproduction end of the chain. Lexicon's digitally implemented "panorama" mode allowed for individual setup adjustments to cater to different loudspeaker/listener geometries. The goal of all of these was to expand the soundstage beyond the stereo loudspeakers, potentially out to ±90°. None

were doing anything that was intended by the recording artists, but all were attempting to reward listeners with an expanded, more enveloping listening experience.

An ambitious attempt to extract the maximum from legacy stereo recordings is Ambiophonics (Glasgal, 2001, 2003). It has gone through several phases of evolution, incorporating binaural techniques as well as complex synthesis of spatial effects to enhance sound delivery. A host of digital "hall" and other artificial reverberation effects have also appeared in audio products; they came to be known as "DSP" effects. Most sounded unpleasantly artificial, but some seemed to have promise.

In the 1990s I supervised the creation of a system called Harman VMAx (Virtual Multi-Axis), that employed the "transaural" crosstalk cancelling method developed by Cooper and Bauck (1989). In addition to cancelling the crosstalk, it could synthesize phantom loudspeakers at specific angles. For example, it could generate a credible illusion of a 5-channel surround system (±30°, 0° and ±90°) °from two loudspeakers placed at ±10°, as on either side of a small TV or computer screen (Toole, 1998). See Figure 3.8.

At two large trade shows, "VMAx Phantom Home Theater" was demonstrated in instantaneous A/B comparisons with a conventional five-loudspeaker implementation, using identical loudspeakers everywhere. The comparison was blind in the sense that the playback form was not announced, just described as A or B, which was randomized. When asked afterwards which system they preferred, about half had no preference. That alone was gratifying, but even more encouraging were the 20% who preferred the phantom theater. Why? The seamlessness of the soundstage and the impressive sense of depth. One was always aware of the five real speakers, whereas the virtual loudspeakers drew less attention to themselves.

A small home theater version was set up, with loudspeakers on each side of a television (±10°) in an acoustically well-damped room, using loudspeakers from a well-designed home-theater-in-a-box system. Listening to stereo recordings through synthesized loudspeakers at ±30°—something we called the "speaker spreader"—was very rewarding, and with certain

classical recordings the stereo soundstage and ambiance were remarkably credible. The impressions of distance and space were impressive—the real loudspeakers and the walls of the room "disappeared"; one responded to cues in the recording. See Figure 3.9.

Although there were simplified computer implementations used by Compaq, and even more simplified audio-receiver implementations in Harman Kardon electronics, VMAx was a commercial disappointment. It was ahead of its time—the digital processing hardware necessary for full implementation was expensive and bulky, and simplified versions were not as persuasive. When properly done, sitting in the sweet spot one can have more rewarding experiences than that in conventional stereo. It clearly depends on the recording, as does stereo. Away from the sweet spot, the experience resembles conventional stereo in the same situation.

Since then, technology and science have advanced, and there is a rich scientific literature on the subject, describing improved crosstalk cancelling filters, such as Bauck and Cooper (1996), Choueiiri (2010), and others adding useful features like head tracking and multiple listeners.

So, let us adjust our mindsets slightly. Let us not think of headphones, or small loudspeakers in close listening situations, as poor substitutes for the traditional audiophile products. Let us view them as legitimate alternatives that, in some important ways, have the potential of being even better. Eliminating or at least attenuating the awareness of loudspeakers and the influence of the listening room are non-trivial factors.

3.4 REPRODUCING THE "PERCEPTUAL ESSENCE" OF LIVE PERFORMANCES

For all the remarkable capabilities of the human binaural hearing system, it is fortunate that in some ways it is easily deceived. If we hear a persuasive clue relating to a perceptual dimension, the brain may "fill in" the missing portion. If we think we may hear something, we may indeed "hear" it. In other words, reproducing the appropriate auditory cues, the "essence", may be enough for a rewarding listening experience.

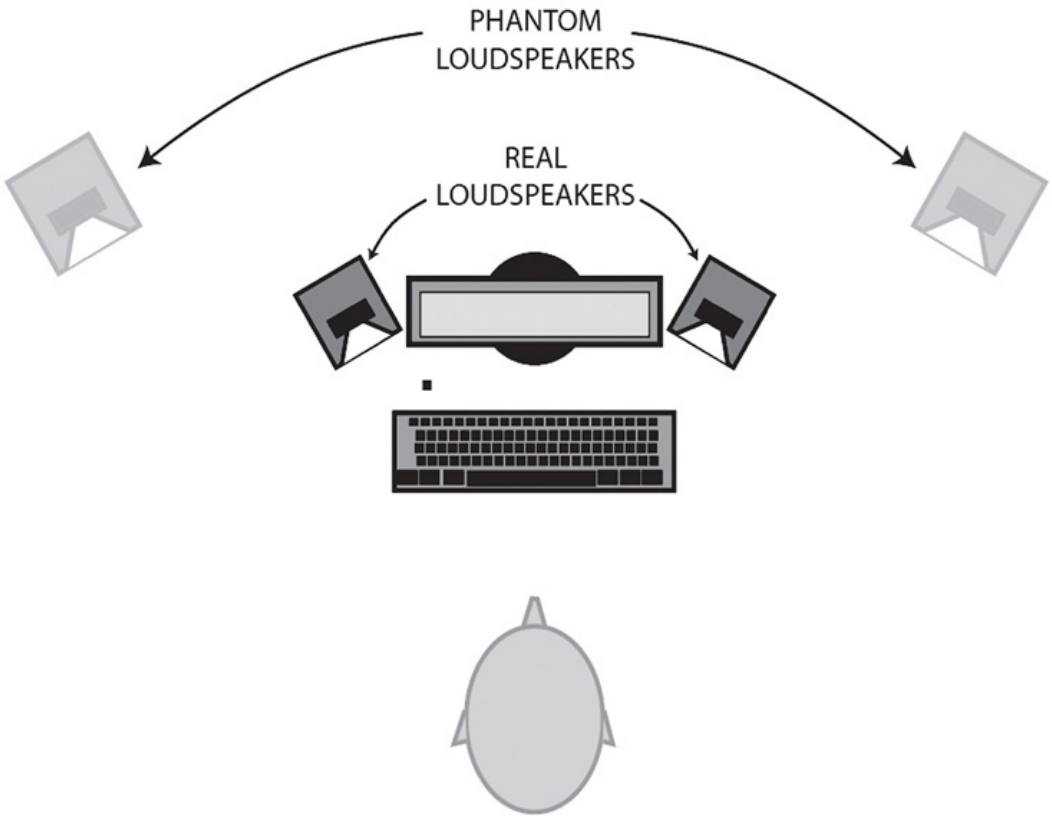


FIGURE 3.9 *With crosstalk cancellation and binaural steering combined, it is possible to position phantom loudspeakers which can then be used to reproduce conventional stereo. The angular separation could be decided by the listener.*

In timbre, an example is the "missing fundamental" bass phenomenon. If the harmonics of a bass note are present, but the fundamental frequency is missing, we are likely to perceive the pitch as that of the missing fundamental. The physically engaging bass energy is missing, but the melodic information is preserved. This is an obvious benefit to small loudspeakers.

The stereophonic illusion of images floating in the space between two loudspeakers is another example. There are no sound sources in the locations of the phantom images, yet the illusion is persuasive for the person in the sweet spot.

In terms of the important sense of envelopment—thinking one is in a large space, as in a concert hall—there are recordings in which this

impression is strongly suggested in 2-channel stereo, but it is more persuasively delivered by multichannel configurations. The essential clue seems to be significantly delayed (>80 ms) early reflections arriving from the side (□60°). These sounds can be captured in the performance venue or electronically synthesized, and they are best delivered by dedicated channels and loudspeakers.

Envelopment is a significant measure of the success of a concert hall. The movie industry recognized the importance of envelopment early on and eagerly adopted multichannel audio when technology made it possible. In addition to being able to deliver off-screen sounds, they wanted audiences in cinemas to get the impression that they were in the venues portrayed on the screen. Concert videos greatly benefit from seeing and hearing the venue and being aware of the surrounding audience.

Nowadays, some spatial embellishments can be created for loudspeaker stereo using the same kinds of production tools that had their origins in binaural synthesis. One can sometimes experience this when listening to the tiny audio systems in laptop computers—sounds may appear to extend beyond the narrow span of the keyboard. Hearing sounds outside the left-to-right span of a pair of stereo loudspeakers may be the result of manipulations in the recordings, having nothing directly to do with the playback hardware or the room.

In summary, all genres of music benefit from an understanding of human perception, and how it can be manipulated—playing games with the hearing system. Recordings should be judged as what they are: artistic creations in their own right, which opens up many opportunities for liking or disliking the result. In all cases loudspeakers are the “messengers”, so in the consumer world they often get the credit or blame.

3.5 RECORDING CONTROL AND MASTERING ROOMS—WHERE “LIVE” PERFORMANCES OCCUR

Figure 3.10 illustrates that in live concert performances the acoustics of the venue are an inseparable part of the experience. There is nothing between

the listener and the performance.

The process for most music recordings and movies is complex, as shown in the center portion of Figure 3.10. It usually begins with the sound from performers in acoustical spaces, sometimes different spaces for different performers, being captured by microphones. Direct pickup signals from electric guitars, basses, keyboards and synthesizers may also be added. In movies, there are Foley effects—real or synthesized sounds that enhance the actions on the screen.

The signals from these sources are manipulated as necessary and stored in individual digital files (tracks) all perfectly synchronized. This can take place over an extended time period, perhaps at multiple venues. At some point they are "mixed": combined, balanced, equalized, compressed and embellished in various ways, with the manipulations being monitored through loudspeakers in a room.

Music recording studios can be costly, stylish places, many of them designed by "name brand" consultants, all suggesting excellence in sound. The investment is also made to attract customers—there is a marketing component. But the era of large expensive recording facilities is fading, as sound mixing and processing programs and computers can perform the functions of traditional massive mixing consoles and outboard processors. More recordings are increasingly being created in modest surroundings, converted bedrooms or garages in homes. Many people now have capabilities that once were the exclusive domain of elaborate recording studio facilities.

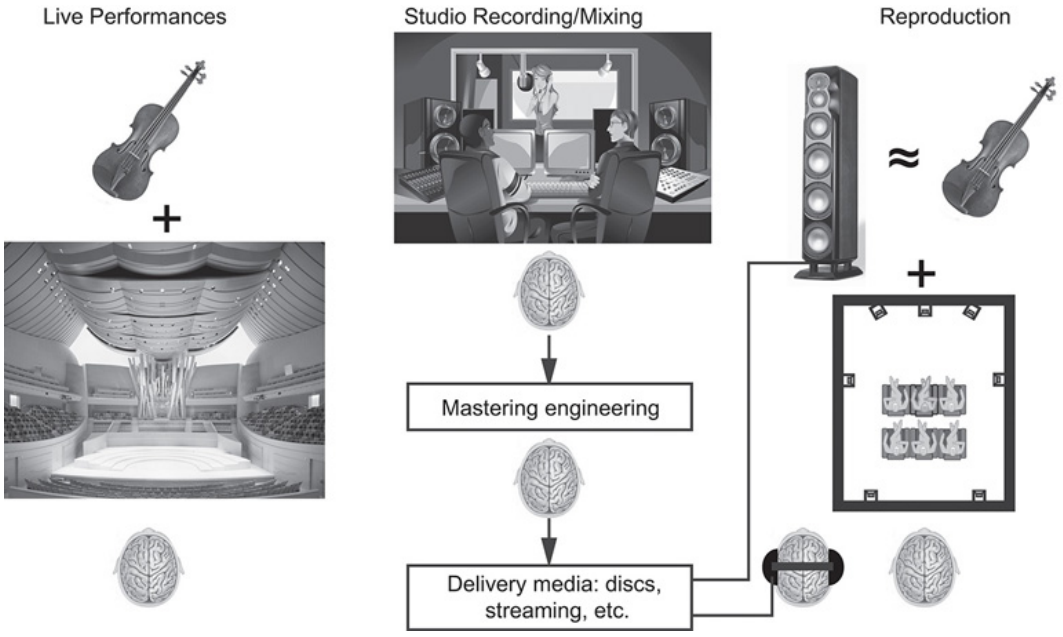


FIGURE 3.10 *A conceptual view of the parallel universes of live performances—sound production—and sound recording and reproduction. The latter process is influenced by human listeners at several points along the way. Disney Hall photograph by Federico Zignani.* ⏎

The internet and AI are aiding this transformation, as live musicians can play together from different locations, simultaneously, or be credibly replaced by samples and artificially created instrumental and orchestral sounds. There are positive and negative consequences of this paradigm shift. But there is no doubt that it is expanding the repertoire of recorded music, changing the business model of the music industry itself and liberating creative instincts that previously had been "damped by dollar signs". From the sound reproduction perspective, there is an elevated need for quality sound reproduction—monitoring—in rooms that may not be acoustically optimized—much like domestic audio.

Professionally designed control rooms embrace almost the entire spectrum of acoustical possibilities. At one extreme is the "non-room" approach that attempts to eliminate all first reflections (Newell, 2003; Walker, 1988; D'Antonio and Konnert, 1984). In the middle is the "live-

end, dead-end" approach in which the front half of the room is acoustically absorbing and the rear end is treated with reflectors and diffusers (Davis and Davis, 1980). At the other extreme is one that has all walls and the ceiling covered with highly effective, deep, diffusers: Blackbird, Studio C, designed by Peter D'Antonio and George Massenburg (see Section 12.5.3). The listening experiences in these rooms cannot be the same, yet they all have, or had, their supporters.

ITU and EBU recommendations. The ITU-R BS.1116-3 (2015), EBU Tech 3276–2nd edition (1998), and EBU Tech 3276-E (2004) documents are often referenced for guidance in setting up systems for sound reproduction. The loudspeaker and room acoustical requirements are identical, yet they have very different purposes.

- The ITU document is intended for blind listening tests evaluating codecs having "impairments so small as to be undetectable without rigorous control of the experimental conditions and appropriate statistical analysis." It is meticulously technical in its objectives, asking the question: is there a perceptible difference between two sounds, and how large is it? This is done in rigorously administered blind tests using listeners who are audiometrically tested, trained and statistically evaluated.
- The EBU documents are intended "for the critical assessment and selection of programme material for inclusion in a sound or television broadcaster's programme output." In distinct contrast, the application of the EBU document is artistically subjective, asking the question of recording engineers: does the program sound good enough to be broadcast?

The ITU listeners look for differences between two sounds, *relative judgments*. The details of the listening circumstances are constant factors, and placing the listeners in a dominant direct sound field may

be advantageous. The EBU listeners are making *absolute judgments* of sound and spatial qualities. The performances of the loudspeakers and room are inseparable parts of the sound they are judging. These listening experiences become the reference for the "translation" of the art. It is the "Creation of the Art" component in the circle of confusion shown in Figure 1.8.

Audio science has progressed, and these recommendations are not up to date. Academics, particularly, rely on standards to validate their results, so it is important that the recommendations are themselves validated. Recommendations are also made for the selection of headphones using the diffuse-field performance target that needs revision (Chapter 15). OR Chapter 3 expands on this discussion and the relevance of these recommendations to sound reproduction in professional and consumer audio domains.

Fazenda et al. (2015) presents a thoughtful survey of the most popular recording studio designs by three well-published acoustical experts associated with the acoustics department at Salford University. In the course of the discussion, they say:

> Another criticism that has been voiced by some is that sound engineers working in a non-environment have the additional difficulty of translating what they hear in a dead space, where they are mixing, to the livelier acoustic of typical domestic listening environments.

So, are they satisfying themselves at the expense of their customers? Perhaps the most telling comment is the conclusion:

> When considering which room design philosophy to follow, it is important to consider your own personal preferences. One of the reasons so many different control-room designs have emerged is because there is not one perfect acoustic solution for a control room. If you have access to other studios, then you should do as architects do before building grand concert halls: do a listening tour to find out which style of control room sounds best to you.

There is no obvious attempt to relate the control room acoustical experience to that of the consumer—the all-important "translation". There is no reported science that includes double-blind subjective evaluations of any of the designs, just opinions from recording engineers who may or may not have damaged hearing (an occupational hazard). It is believed by many in pro audio that the choice of monitor loudspeakers is a personal thing. So, we seem to be in a situation where individual recording engineers choose loudspeakers and control room designs on the basis of preference, without regard for existing science, or concerns about validating whether "translation" is possible. None of the control-room designs being discussed is likely to occur in homes, or even custom home theaters. This is not rational.

Augspurger (2013), a well-known acoustical consultant, provides an insightful review of acoustical practice in rooms used for tracking, mixing and mastering, observing that:

> After 1990 or so, although professional recording engineers had reached a consensus regarding the characteristics of a good mixing room, fewer and fewer such rooms were being built. The Pro Tools digital workstation had become the accepted standard for tracking, processing, and mixing recorded music. Almost every music composer and producer acquired a Pro Tools setup and proceeded to use it in the nearest convenient location. A spare bedroom became a professional music production room. If we include music composed for television and movies, the bulk of music produced in the U.S. today probably comes from residential studios. A major challenge for studio designers is how to make a small room acoustically acceptable for stereo monitoring and mixing.

If being "acoustically acceptable" means experiencing timbrally neutral reproduced sound, the formula must begin with properly designed loudspeakers, followed with some of the recommendations in this book. This is territory not familiar to everybody in professional audio.

There are several reasons why recording engineers might wish to be in a predominantly direct sound field. They obviously need to be able to hear the results of their manipulations, including hearing the acoustics of the

recording venue with minimal distraction by the acoustics of the listening space. This is part of the artistic creative process. Low reverberation times, absorbing or deflecting early reflections and near-field monitoring all contributed positively to achieving this condition. Not discussed is the embarrassing fact that several of the very popular large monitor loudspeakers had such poor, colored, off-axis performance that it was best not to be heard at all. Several are shown in the following sections. Sound quality was improved by absorbing these corrupted sounds, reducing the sound field to dominantly direct sound. There were multiple motivations for controlling reflections.

Mastering engineering finishes the recording process, and it is conducted by a different person listening to different loudspeakers in a different acoustical space. This person further manipulates the recording to meet expectations of sound and spatial qualities and also to be appropriate for an intended audience or delivery format. It is reasonable that this person should listen in a situation that resembles that of the average consumer, employing loudspeakers of similar design in a room having similar acoustics. According to Hinksman (2021), mastering engineers might have been relegated to non-descript rooms, although a few lucky ones got acoustically treated spaces. Others set up operation in their homes, a normal domestic space, an acoustically treated domestic space, or a modified garage. There is the belief that intimate familiarity with the room is more important than the nature of the room itself. Again, there are no standards.

Compression is usually part of mastering engineering, often referred to as an artistic "glue" providing unity to the performance. The dynamic range is reduced, attenuating the loud portions and amplifying the softer portions. Moore (2021) says that it imposes "a uniform control over dynamics, which imparts a cohesiveness to program material. Depending on the style of the compressor, this may include the introduction of subtle uniform harmonic distortion, colouration and a sense of rhythmic movement and groove."

The mastering engineer can be a significant contributor to how the final recording sounds—the last human involved in the creative process. However, things change, and now streaming is the prevailing delivery

medium for music and movies and it requires attention to average loudness levels of program, as well as other factors. Mastering in some applications has apparently become so routine that much of it is relegated to automated processors. Adding to the change is the vision by some that AI can do even more. This is a major development that will affect mastering engineering as a profession (Elliot and Chon, 2022). One hopes that humans continue to contribute to the art.

3.5.1 "Monitor" Loudspeakers Are Not Necessarily "Good" Loudspeakers

As much as we respect the artistic skills of professional recording engineers, it is unfortunate that they have not always had the best tools to work with. Figures 1.4 and 1.5 illustrate the important role of loudspeakers in the record/reproduction cycle. Figure 2.7 shows that professional monitor loudspeakers were not leading the way to neutral sound reproduction from 1980 to 1985 onward, musically not all that long ago.

Figure 3.11 is persuasive evidence that for important decades in music recording professional recording engineers were exposed to a wide range of timbral signatures from their monitor loudspeakers. The important direct sound (on axis) was extremely variable, and some of the curves were so irregular that even equalization of the period would not have been effective. The frequency-dependent directivity inconsistencies exhibited by some (e.g. d, e, f, g, h, i) meant that off-axis room reflections would convey distorted timbral messages, competing with the direct sound. These were not reliable references of sound quality, and a few were simply bad loudspeakers. Yet, they were what was available, and were widely used. The belief that professional "monitor" loudspeakers are inherently superior to those in the consumer world is not supported by the evidence. They may play loud and long without breaking, but that is all. Of this collection of products, only (j) shows evidence of being designed to deliver relatively neutral sound quality in normal listening spaces. For interest compare it to Figure 1.3(f), the loudspeaker that was rated highest in my first listening tests 22 years earlier. An article, Toole (1979), explains my feelings at the time: monitor

loudspeakers need to meet “hi-fi” standards. I’m sure it was not taken seriously.

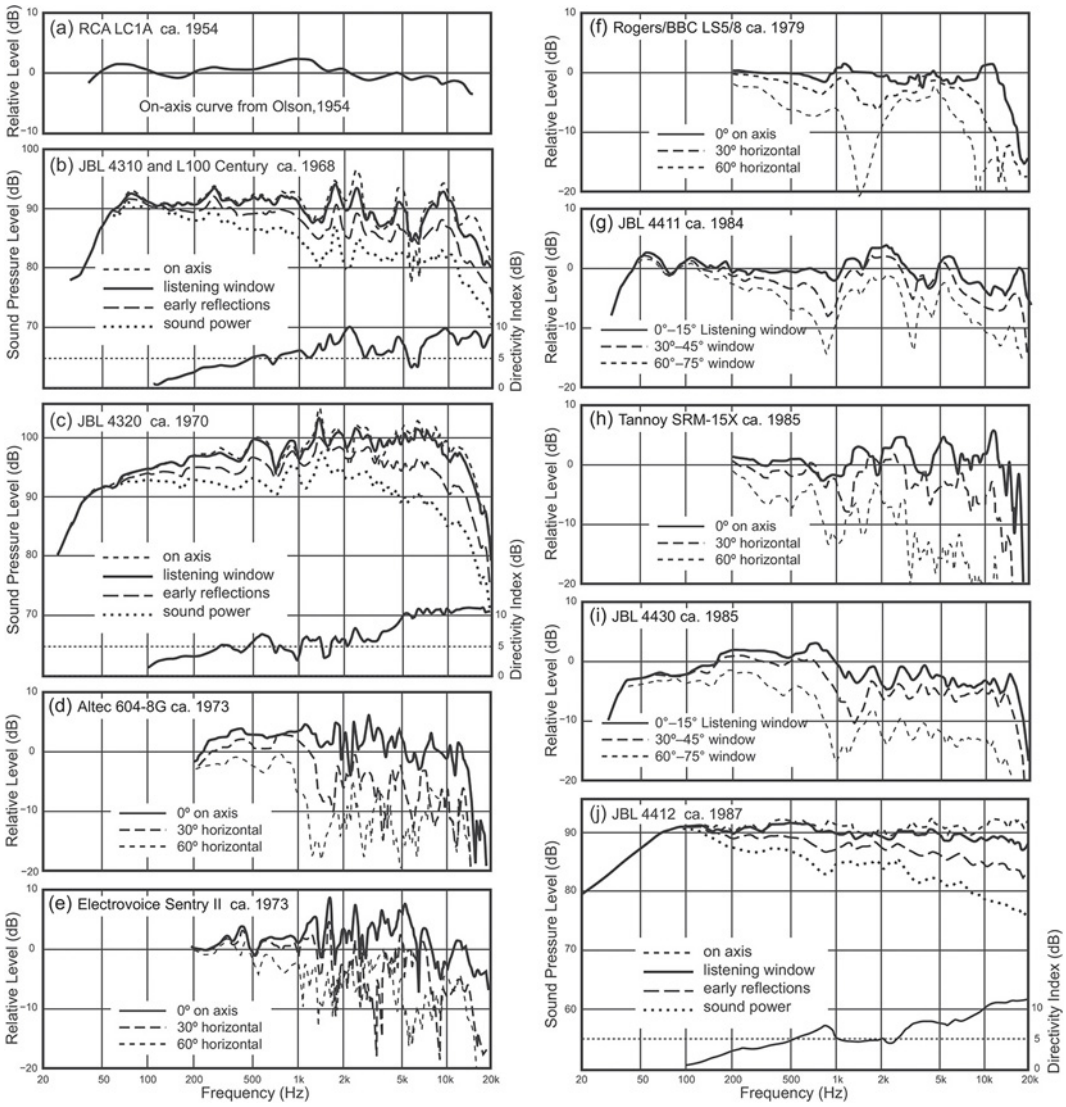


FIGURE 3.11 *A sample of medium and large professional monitor loudspeakers I tested over the years at the NRCC and at Harman. The exception is the historic RCA LC1A, which is an obviously smoothed version from Olson et al. (1954). Some of the curves are truncated at 200 Hz because the chamber had not yet been calibrated.* ⏎

Main monitor loudspeakers are often built into the front wall of the control room and are expected to be able to play very loud, often at hearing damaging levels, without breaking—expensive "dead air" is to be avoided. Performance trade-offs have been necessary, so, as can be seen in Figure 3.11, some popular examples of these loudspeakers were quite colored, with very poor off-axis behavior. Recording engineers move around among studios and the sound in each one is different, so to provide continuity small "near-field" monitors, sitting on the meter bridge of consoles or similarly located for "close listening", are commonly used.

Today, loudspeakers are much improved, and there are new approaches to optimizing bass performance in small rooms. These are advantages for the new generation of converted domestic spaces being used as recording venues and control rooms. A concern over hearing loss has also resulted in some moderation in monitoring sound levels.

3.5.2 An Acoustic Digression in Pro Audio

Around 1960 an inexpensive 4.5-inch (114 mm) transducer, widely used at the time in TVs and table radios, was put in a small box and promoted as a reference loudspeaker, for use by recording engineers to assess whether their mixes would "translate" into the "real world". Figure 3.12 shows its performance. Not surprisingly it was not very good, and not consistent—the three samples differed significantly. The Auratone 5C was widely used. This made sense 60+ years ago. Since then, though, much has changed in the audio industry, including the "real world".

The recording industry, as a whole, is very professional, not easily influenced by things that don't make sense. However, there have been exceptions. The distorted sonic signature of the Auratone 5C became so entrenched that it reappeared years later in the Yamaha NS-10M Pro, as shown in Figure 3.12. Its on-axis frequency response was a smoothed version of an Auratone, but with more bass. It became a popular near-field monitor in recording control rooms around the world. The origin of the NS-10M is a separate story and it had nothing to do with the recording industry. It was a flawed consumer product, but because of its acceptance by

recording engineers, Yamaha built a sturdier Pro version. Discontinued years ago, they are still found in many recording control rooms. Imitations are now being built.

Around the same time, a "time-aligned" version of the large Altec 604–8G (Figure 3.11(d)) came to the market as the UREI 811, 812 and 813, with different numbers of woofers. It was widely used. Figure 3.13 shows how it perpetuated the basic sonic signature of the Auratone 5C with more bass, and at much higher sound levels.

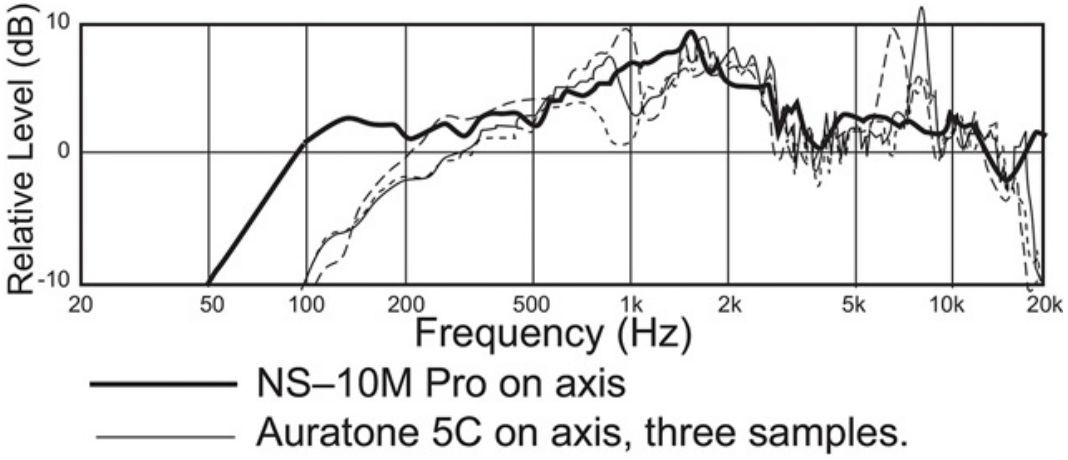


FIGURE 3.12 *On-axis frequency responses of the Auratone 5C and Yamaha NS-10M Pro loudspeakers.* ⏎

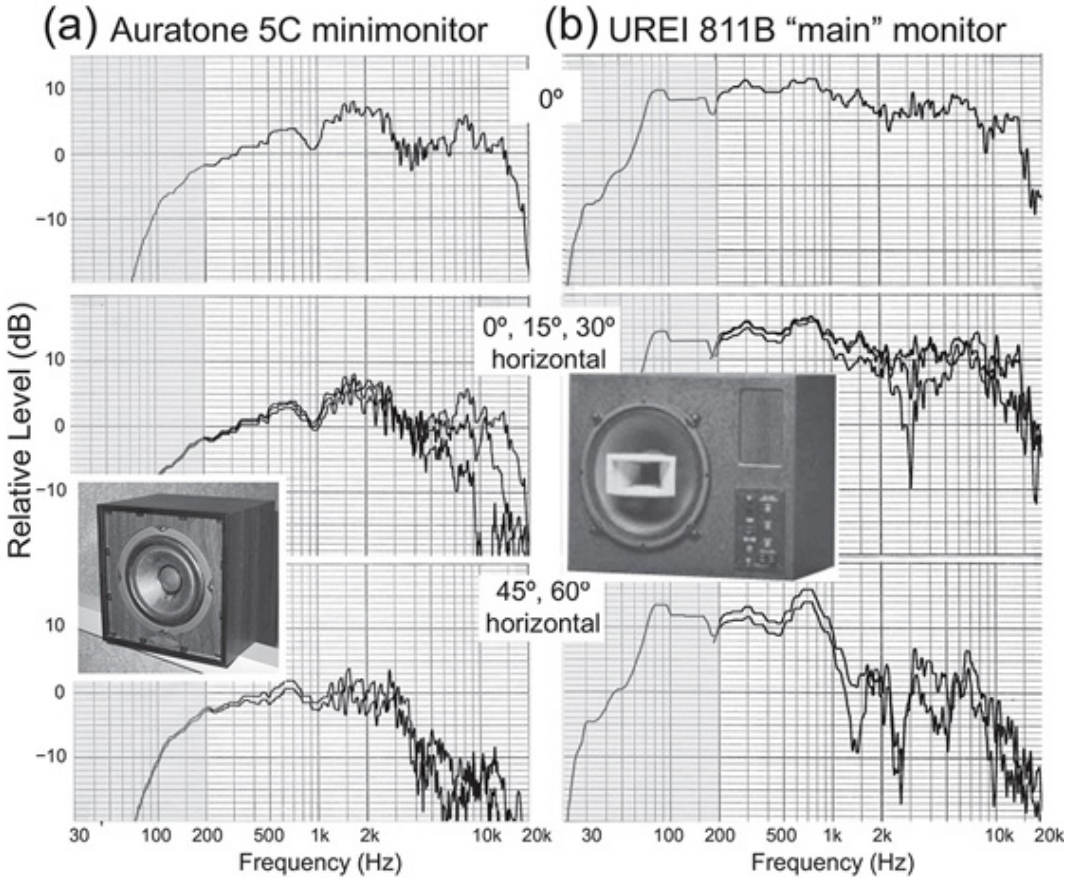


FIGURE 3.13 *A comparison of anechoic data on the 4.5-inch Auratone 5C and a 15-inch woofer with concentric horn UREI 811B, showing fundamental similarities in middle and high-frequency performance.* ⏎

Figure 3.13 conveys a persuasive message. Considering the huge difference in their sizes and acoustical output capabilities, it is remarkable how similar these on- and off-axis anechoic measurements are. It is difficult to justify the use of these seriously flawed loudspeakers by professionals in in the 1980s, much less later. In a series of monitor loudspeaker evaluations I did with the Canadian Broadcasting Corporation, the UREI 811B was subjectively rated for sound quality in double-blind tests, and it received the lowest rating. Listeners included numerous recording engineers and producers (Toole, 1985, 1986). Harman acquired UREI in 1983 for its

electronic products. Production of loudspeakers was terminated before I arrived in 1991.

Perpetuating the trend, some of these old bad-sounding loudspeakers now have modern versions, finding justification in claims that they accentuate certain sonic features during the tracking and mixing processes. This is subjective audio at work. Does it matter? It does if these "distinctive" sounding monitors are used to evaluate the final product. Some mastering engineers have commented that they can recognize when aberrant loudspeakers have been used in the mixing process and compensate for them with equalization. So, it matters to somebody.

Because near-field monitors deliver dominant direct sound, the right kind of equalization can transform a neutral loudspeaker into countless different "loudspeakers" at the push of a button, thereby reducing acoustic clutter around the recording engineer. Bass performance in amplitude and time is customizable, because woofers are minimum-phase devices, but the sound is dominated by room resonances that need separate attention (Chapter 14). A recording engineer who chooses a loudspeaker because the sound is appealing should find out what causes it and build that quality into recordings. Using that loudspeaker as a monitor ensures that only those in the control room will experience the effect—unless it happens to be a neutral loudspeaker. Not believing in or knowing anechoic measured loudspeaker performance is a professional handicap.

Monitoring with timbrally neutral loudspeakers is defensible logic—think of them as the average of all possible "bad" loudspeakers—a transparent "window". Real-world evidence supports this. There is no standard for bad sound, but we now know that there is a common goal for high sound quality. A problem for everyone is that not all manufacturers are forthcoming with meaningful anechoic performance data on their products, so identifying a potentially neutral loudspeaker is more difficult than it needs to be.

Now, basically neutral-sounding professional monitor loudspeakers of all sizes are available. Some find their way into high-quality home systems, and good domestic loudspeakers make their way into professional

environments. After all, a good loudspeaker is a good loudspeaker, so long as it can play loud enough and long enough for the purpose. There is hope that in future the circle of confusion (Figure 1.8) will be less of an issue.

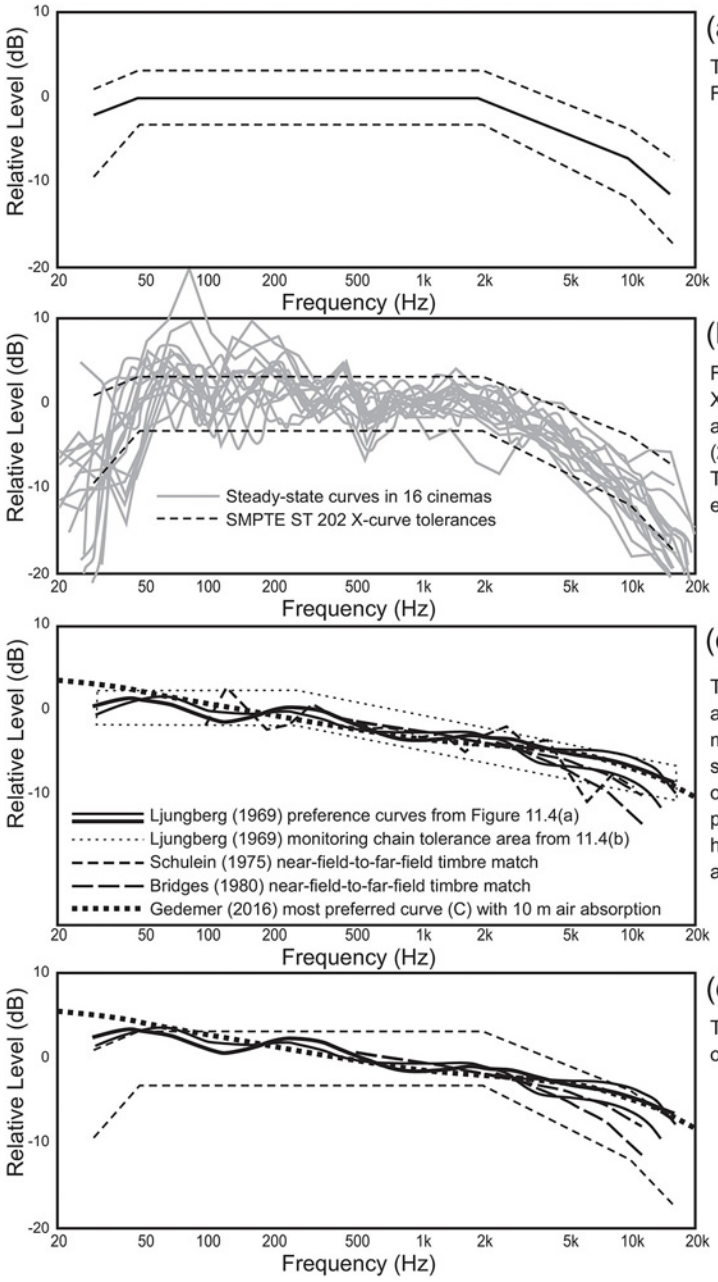
So, to summarize this section, the process of creating recordings is a complex operation, involving human listeners at several stages. These listeners create the art, but through their opinions they have also influenced the technology—the loudspeakers and room acoustic designs within which the art is crafted. Unbiased listening evaluations are scarce. The stage has now been reached where we can be much more *objective* about the technology, leaving *subjectivity* to the art created with it. Section 12.5 discusses what might be needed for "translation".

3.6 THE CLOSED LOOP OF CINEMA SOUND

The movie industry is an island in the total audio industry. From the outset, the close relationship between the content creators and cinemas was an advantage. The intention was that movie makers and cinema owners would follow the same the industry standards for production and playback loudspeakers and venues. This would ensure consistency in the delivered sound, avoiding the circle of confusion described in Figure 1.8. It was an excellent concept, and standards and calibration methods were established. Unfortunately, the common acoustical measurements of the period were primitive: 1/3- or octave-band analysis of pink noise, yielding steady-state room curves which are not reliable indicators of sound quality at all frequencies. Compounding the problem was the target curve, the so-called X curve, which was an unfortunate choice. There was published research at the time (Ljungberg, 1969) in the *Journal of the Society of Motion Picture and Television Engineers* (SMPTE) providing better guidance than was employed in defining the X curve that became the standard. Later research confirmed it (Figure 3.14).

The result is that the spectral balance target of cinema sound is different from that which is the expectation in the rest of the audio industry, consumer and professional. It is not neutral, requiring spectral

compensation by the recording engineers to arrive at a natural timbral balance. A serious added complication is that the system calibration method, using room curves, yields sound quality that varies among venues. It has been argued that despite these concerns the system has functioned better than if there had been no standards, which is likely true. However, science and technology moved on long ago, and industry standards have not changed.



(a)

The X-curve and its tolerances. From: SMPTE ST 202 (2010).

(b)

Full-bandwidth measurements in16 X-curve calibrated cinemas. The curves are from Figures 16 and 17 in Toole (2015) using data from SMPTE TC-25CSS (2014), and Newell et al. (2010).

(c)

The results of three timbre-preference and two near-field-to-far-field timbre matching experiments. All are shown superimposed on the tolerance area for octave-band filtered measurements published by Ljungberg (1969) which he proposed August 19, 1967 - 58 years ago!

(d)

The curves shown in (c) superimposed on X-curve tolerances.

FIGURE 3.14 *An illustration of several relevant measurements related to cinema sound.* ⏎

Figure 3.14(a) shows the cinema sound X curve, the steady-state room curve measured at five locations around 2/3 distance from the screen, using 1/3-octave band analysis of pink noise. It departs from such curves measured in most audio systems because there is no overall downward tilt, the normal result obtained when forward-firing loudspeakers radiating flat, direct sound are measured in reflective spaces. The downward tilt arises because woofers become progressively more directional as frequencies rise. Cone and dome systems tend to continue to become more directional at mid and high frequencies, but horns can be designed for constant directivity over most of their operating range. The other feature to note in these straight-line target curve segments is the sharp knee at 2 kHz, something that is possible to be audible as a coloration. This is an unnecessary fault.

Figure 3.14(b) suggests that most of those cinema calibrators noticed the deficient bass and were unwilling to turn the bass down to meet X-curve requirements. Strangely, the very high frequencies were often allowed to drop below the target. There are several possible explanations, most having to do with the large boosts often required at very high frequencies to overcome screen loss, air attenuation and fading output from the loudspeaker drivers, especially older designs.

Figure 3.14(c) shows the results of subjective tests of preferred sound quality in large venues. The results are extraordinary in that seven room curves arrived at by four experimenters over a 47-year time span cluster closely around each other over most of the frequency range. Most fall within the 4 dB tolerance window proposed by Ljungberg in 1967 and published in 1969. They all are close approximations of what would be measured from loudspeakers radiating a flattish, direct sound. This is a direct parallel to subjective preferences in domestic listening spaces.

Figure 3.14(d) shows a comparison between the findings shown in (c) and the X-curve tolerances from (a). Insufficient bass and a sharp knee at 2 kHz are obvious concerns.

Now, movies are enjoyed outside cinemas more often than in them, being delivered to customers in their televisions and home theaters as well as planes, trains and all manner of personal devices. All of these sound delivery systems are designed to meet different performance objectives. The closed loop is broken, and the circle of confusion exists.

Figure 3.15 is important because it shows that the key element in subjective preferences is the direct, on-axis sound from the loudspeaker. If the off-axis frequency-dependent directivity is well behaved, the resulting steady-state room curve is a tilted line over most of the frequency range as the sound output ranges from omnidirectional in the low bass, progressively becoming more directional at higher frequencies; the radiated sound power is gradually reduced. The bass rise is also influenced by the reflectivity and/or reverberation time of the room at low frequencies, and the high-frequency roll-off above about 1 kHz is determined by air attenuation related to listening distance.

The message in Figure 3.15 is that in both large and small rooms, listeners have preferred a flattish, direct sound. In room curves of well-designed loudspeakers, this results in relatively linear room curves with downward tilts. In large rooms the tilt is more pronounced because bass is elevated by higher reverberation times and treble is lowered by air attenuation because listeners are farther from the loudspeakers.

The problem is that these room curves get interpreted as “targets” for equalization of loudspeakers of unknown performance. What is measured is dominated by the off-axis performance of the loudspeakers—the reflected sound—which is influenced by the acoustical properties of the venue. Since neither the loudspeakers nor the rooms are tightly specified, the measured room curves are “results”, not targets. They can only be trusted for evaluating sound quality if the detailed anechoic performance of the loudspeakers is known and meets certain standards, and the acoustical nature of the venue is closely standardized. This could be the case in cinemas if the playback (B-chain) components were specified in sufficient detail, but they are not, and it is a worldwide industry with numerous independent owners and suppliers.

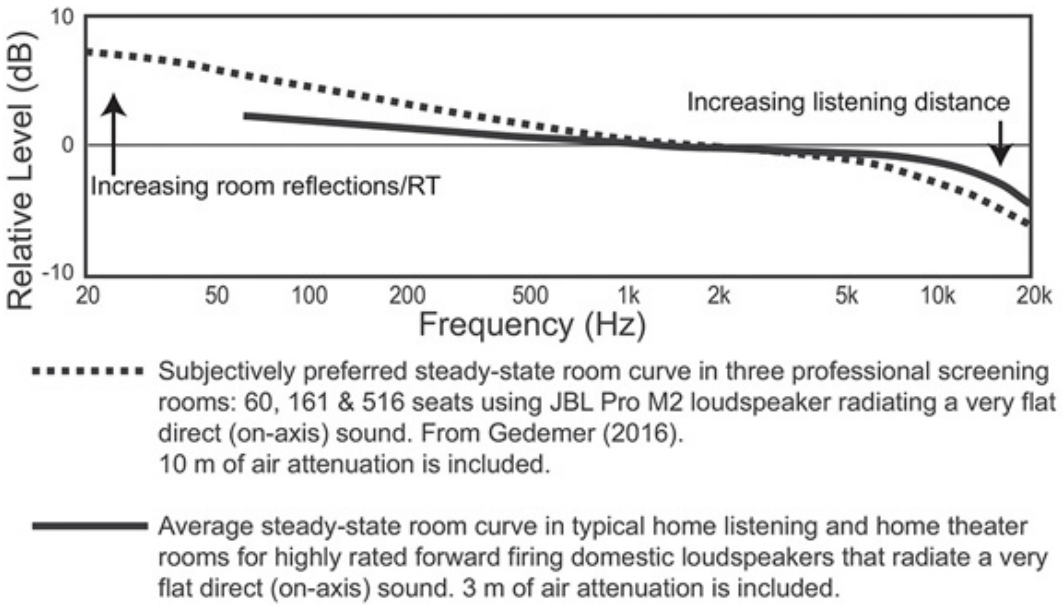


FIGURE 3.15 *A comparison of steady-state room curves in professional cinema venues and homes, using loudspeakers that have been subjectively highly rated in double-blind evaluations. All loudspeakers had very flat and smooth on-axis frequency responses and smoothly changing off-axis responses.* ⏎

As shown in Figure 3.14(b), some cinema system calibrators could hear problems and deviated from the standard, boosting the bass. Recording engineers also made changes, sometimes having two system settings, one for the calibrators and one to satisfy themselves. Conveniently, many of the "unofficial" modifications to the calibration X curve allowed recording engineers and audiences to hear sounds more like the rest of the audio industry. This is an unsatisfactory situation, and eventually a committee was set up within the SMPTE standards group to investigate the situation. I was invited to be on that committee. Measurements were made in numerous calibrated cinemas and mixing facilities. It was concluded that there were indeed problems, the reasons were understood, and it was time to revisit the standard using modern measurement methods and a revised performance target (SMPTE TC-25CSS (2014). One movie "mix" should be compatible with all audiences, in cinemas and homes, using loudspeakers, headphones or VR devices. Repurposing should not be required.

Now there is a new factor: streaming services are producing their own movies and series. Hollywood and its derivatives have competition, and the competition may have its own ideas of what constitutes “good sound”. Its customers are not bound to cinemas.

The arrival of immersive sound is an opportunity to change things because cinemas will need some equipment changes. There are signs that in calibrating the new Dolby Atmos facilities, many of which are smaller new constructions dedicated to music production, the target curve recommendation has been modified from the original X curve. Section 8.4.3 discusses the new recommendations in Dolby (2024).

Fortunately, when a good movie begins disbelief is suspended, attention is diverted from considerations of sound quality, and with enough sound level and bass, audiences are generally pleased. Those wishing to know more technical detail about the movie sound situation will find it discussed in considerable detail, with graphs and references, in chapter 11 of the third edition of this book, which is reproduced at OR Chapter 3.

3.7 SUMMARY

The lack of consistent technical standards in the music recording and movie industries is a problem for both artists and consumers of the art. Art is created with the best of intentions that it will translate across production facilities during the creation process, and then translate to whatever playback devices or circumstances prevail. Flawless translation is impossible. Dissatisfied listeners often choose to blame the last device in the chain of events: the loudspeaker, and possibly the room. This may or may not be true. Fortunately, the movies and music can still be pleasurable, such is the adaptability of human listeners, but the situation can definitely be improved, as discussed in Section 12.5.

REFERENCES

Augspurger, G. L. (2013). “Control Room Design: The Monitoring Environment”, *Acoustics Today*, April, pp. 15–20. https://acousticstoday.org/wp-content/uploads/2017/10/Article_2of4_from_ATCODK_9_2.pdf ⏎

Bauck, J. and Cooper, D. (1996). “Generalized Transaural Stereo and Applications”, *J. Audio Eng. Soc*., vol. 44, no. 9, pp. 689–705. ⏎

Bregman, A. S. (1999). *Auditory Scene Analysis, the Perceptual Organization of Sound*, MIT Press. ⏎

Bridges, S. (1980). “Effect of Direct Sound on Perceived Frequency Response of a Sound System”, *Audio Eng. Soc., 66th Convention*, Preprint 1644.

Carver, R. (1982). “Sonic Holography”, *Audio*, vol. 66, March. ⏎

Choueiiri, E. (2010). “Optimal Crosstalk Cancellation for Binaural Audio with Two Loudspeakers”. https://3d3a.princeton.edu/publications/optimal-crosstalk-cancellation-binaural-audio-two-loudspeakers ⏎

Cooper, D. and Bauck, J. (1989). “Prospects for Transaural Recording”, *J. Audio Eng. Soc*., vol. 37, pp. 3–19. ⏎

D’Antonio, P. and Konnert, J. H. (1984). “The RFZ/RPG Approach to Control-room Monitoring”, *Audio Eng. Soc. 76th Convention*, Preprint #2157. ⏎

Davis, D. and Davis, C. (1980). “The LEDE Concept for Control of Acoustics and Psychoacoustic Parameters in Recording Control Rooms”, *J. Audio Eng. Soc*., vol. 28, pp. 585–595. ⏎

Dolby (2024). “Best Practices for Dolby Atmos Music Studios”, September 5. https://dolby.my.salesforce.com/sfc/p/#700000009YuG/a/4u000000lFIV/YHgXgrJVKAcelcXkD.DnlRUBZU1CqEMGDKweLRpuu98. Accessed September 2024. ⏎

EBU Tech 3276–2nd edition (1998). “Listening Conditions for the Assessment of Sound Programme Material: Monophonic and Two–channel Stereophonic”, *European Broadcasting Union*, Geneva. ⏎

EBU Tech 3276-E (2004) Supplement 1, Revised May 2004 “Multichannel Sound”. ⏎

Elliot, M. and Chon, S. H. (2022). “A Comparative Study of Music Mastered by Human Engineer and Automated Services”, *J. Audio Eng. Soc*., vol. 70, no. 8, pp. 764–778. ⏎

Fazenda, B., Angus, J. and Cox, T. (2015). “The SOS Guide to Control Room Design”. www.soundonsound.com/techniques/sos-guide-control-room-design. Accessed October 2024. ⏎

Gedemer, L. (2016). “Subjective Listening Tests for Preferred Room Response in Cinemas—Part 2: Preference Test Results”, *Audio Eng. Soc., 140th Convention*.

Glasgal, R. (2001). “Ambiophonics”, *111th Convention, Audio Eng. Soc.*, Preprint 5426. ⏎

Glasgal, R. (2003). “Surround Ambiophonic Recording and Reproduction”, *24th International Conference, Audio Eng. Soc*. ⏎

Griesinger, D. (2019). “Accurate Reproduction of Binaural Recordings through Individual Headphone Equalization and Time Domain Crosstalk Cancellation”, *147th Convention, Audio Eng. Soc*., Paper 10239. ⏎

Hinksman, A. (2021). “The Creative Mastering Studio”, Chapter 1 in Braddock, J. P., Hepworth-Sawyer, R., Hodgson, J., Shelvock, M. and Toulson, R., editors, *Mastering in Music*, Focal Press. ⏎

ITU-R BS.1116-3 (2015). “Methods for the Subjective Assessment of Small Impairments in Audio Systems”, *International Telecommunication Union*, Geneva. ⏎

Ljungberg, L. (1969). “Standardized Sound Reproduction in Cinemas and Control Rooms”, *Journal of the SMPTE*, vol. 78, pp. 1046–1053. ⏎

Moore, A. (2021). “Towards a Definition of Compression Glue in Mastering”, Chapter 3 in Braddock, J. P., Hepworth-Sawyer, R., Hodgson, J., Shelvock, M. and Toulson, R., editors, *Mastering in Music*, Focal Press. ⏎

Morrison, J. (2009). “Acoustic, Visual and Aural Space—The Quest of Virtual Reality in Musical Reproduction”, in *Explorations in Musical Ecology (EME)*, Hampton Press, pp. 81–98. ⏎

Newell, P. (2003). *Recording Studio Design*, Focal Press, Oxford. ⏎
Newell, P., Holland, K., Torres-Guijarro, S., Castro, S. and Valdigem, E. (2010). "Cinema Sound: A New Look at Old Concepts", *Proc. Inst. Acoust*., vol. 32, pt. 5.
Olive, S. E. and Toole, F. E. (1989a). "The Evaluation of Microphones—Part One: Measurements", *87th Convention, Audio Eng. Soc*., Preprint 2837. ⏎
Olson, H. F., Preston, J. and May, E. (1954). "Recent Developments in Direct-Radiator High-Fidelity Loudspeakers", *J. Audio Eng. Soc*., vol. 2, pp. 219–227. ⏎
Rayburn, R. (2011). *Eargle's The Microphone Book*, 3rd edition, Routledge. ⏎
Schulein, R. B. (1975). "In Situ Measurement and Equalization of Sound Reproduction Systems", *J. Audio Eng. Soc*., vol. 23, pp. 178–186.
Shaw, E. A. G. (1966). "Earcanal Pressure Generated by a Free Sound Field", *J. Acoust. Soc. Am*. vol. 39, pp. 465–470. ⏎
Shaw, E. A. G. (1974). "Transformation of Sound Pressure from the Free Field to the Eardrum in the Horizontal Plane", *J. Acoust. Soc. Am*., vol. 56, pp. 1848–1861. ⏎
SMPTE TC-25CSS (2014). "B-Chain Frequency and Temporal Response Analysis of Theaters and Dubbing Stages", *Soc. Motion Picture and Television Eng*. ⏎
SMPTE ST 202 (2010). "For Motion Pictures—Dubbing Stages (Mixing Rooms), Screening Rooms and Indoor Theaters—B-Chain Electroacoustic Response" (2010).
Steinberg, J. C. and Snow, W. B. (1934). "Auditory Perspective—Physical Factors", *Electrical Engineering*, vol. 33, pp. 12–17. ⏎
Toole, F. E. (1970). "In-head Localization of Acoustic Images", *J. Acoust. Soc. Am*., vol. 48, pp. 943–949. ⏎
Toole, F. E. (1976). "Measuring Headphones", *AudioScene Canada*, vol. 13, Aug., pp. 63–65 ⏎
Toole, F. E. (1979). "Hi Fidelity in the Control Room—Why Not", *dB The Sound Engineering Magazine*, February, pp. 30–33. ⏎

Toole, F. E. (1984). “The Acoustics and Psychoacoustics of Headphones”, *Audio Eng. Soc. 2nd International Conference*, Paper C1006. ⏎

Toole, F. E. (1985). “Subjective Measurements of Loudspeaker Sound Quality and Listener Preferences”, *J. Audio Eng. Soc*., 33. pp. 2–31. ⏎

Toole, F. E. (1986). “Loudspeaker Measurements and Their Relationship to Listener Preferences”, *J. Audio Eng. Soc*., vol. 34, pt. 1, pp. 227–235, pt. 2, pp. 323–348. ⏎

Toole, F. E. (1998). “Direction and Space—The Final Frontiers. Multichannel and Binaural Audio Systems for Music, Movies and Games”. www.harman.com/documents/HowManyChannels_0.pdf ⏎

Walker, R. (1998). “A Controlled-reflection Listening Room for Multi-Channel Sound”, *104th Convention, Audio Eng. Soc*., Preprint 4645. ⏎

Woodcock, J., Davies, W. and Cox, T. (2019). “Influence of Visual Stimuli on Perceptual Attributes of Spatial Audio”, *J. Audio Eng. Soc*., vol. 67, no. 7/8, pp. 557–567. ⏎

CHAPTE
R 4

Sound Reproduction: Part One—Mono to Stereo

Floyd Toole

DOI: 10.4324/9781003477495-5

4.1 THE PRINCIPAL PERCEPTUAL VARIABLES OF REPRODUCED SOUND

What are we trying to reproduce, to imitate? Some say that the live acoustical experience must be the ultimate standard, but typical recording techniques do not capture the performance as it is heard in the audience. Classical music is a blend of sounds picked up within and above the orchestra and other sounds collected back in the hall. Popular music is abstract art, assembled from many pieces (tracks), manipulated in recording control and mastering rooms, where the original "performances" occur. None of this "pure", and that is not a problem because it is art, it is entertainment, and most of it is successful but sometimes it disappoints.

At issue here is whether the record/reproduction system has the ability to capture and reproduce the *principal perceptual variables* that contribute to enjoyable listening experiences. As explained in Section 3.5, the final judgments of master recordings of all kinds are made in small rooms listening to loudspeakers, so for consumers there is a real possibility of hearing something close to the original "performance".

We cannot reconstruct specific sound fields of individual recording control rooms, much less those of concert hall performances. Instead, we deliver sounds that give listeners the perceived impressions of being in different sound fields. The following are what I think might be the *perceptual essences* alluded to in Section 3.4:

- *Sound quality*: This is unrelated to the number of channels, but obviously all channels should deliver satisfying sounds. Do the voices and musical instruments sound natural, real, or are they colored or distorted? Sometimes timbral changes are made for artistic purposes, in which case the listener cannot know what was intended. In movies there are many sounds that have no parallels in reality; they are special effects. Whatever was done in the recording control room or movie dubbing stage should translate—be reproduced in its essence—for listeners in groups or alone. Consumers need to trust that the playback apparatus is a neutral, transparent communicator of what was recorded. At the creative end of the process, there needs to be trust that consumers have a reasonable chance of hearing the art that is being crafted. This is the basis of translation, and sound quality is where it starts. In the end, many consumers use tone controls or broadband equalization to tailor the sound to their liking, but a lack of coloring resonances in neutral loudspeakers is a good starting point.
- *Direction*: the ability to localize sound sources, beginning with a front soundstage to emulate the common paradigm of live concerts and movies. Specific sounds from the sides and rear are good for special effects in movies and can be used to create the illusion of being within a group of musicians, something not universally appreciated, or of

being in the audience. More channels permit more discrete localizations but, as we have learned from stereo, *for listeners in the right locations*, it is possible to create phantom images, called objects, between selected loudspeakers in immersive systems.

- *Distance*: a component of a recording delivered to only a single loudspeaker is perceived at the distance of that loudspeaker. If simulated or real reflections of that sound are added from other channels, or if binaural processing is employed (see Section 3.3.4) it is possible to create the illusion of greater distance. It is exciting to perceive sounds originating outside the boundaries of a room or car. An interesting observation from many double-blind listening tests is that the highest-rated loudspeakers tended to "disappear"—not completely, obviously, but in the sense that they did not draw attention to themselves behind the screen. Even in monophonic listening there seemed to be a greater sense of distance. My own observation about surround loudspeakers in multichannel systems is that the more neutral the loudspeakers, the more persuasive the spatial impressions. Recorded spaces are likely to be larger than typical domestic listening rooms.
- *Spaciousness or spatial impression*: perceptions associated with listening in a space, especially a large space. It has two principal perceptual components, ASW and LEV:
  - *Apparent source width* (ASW) is a measure of perceived broadening of a sound image whose location is defined by direct sound. In live performances it is the auditory illusion of a sound source that is wider than the visible sources; this is considered to be a strongly positive attribute of a concert hall. Some audiophiles think of it as "air" around instruments. Other audiophiles have come to think that pinpoint localizations are a measure of excellence—opposite points of view that lead to long internet debates about optimum loudspeaker directivity.

- *Listener envelopment* (LEV) is a sense of being in a large space. The direct sound from a source is followed by much-delayed reflections from room boundaries, especially the side walls. The delayed versions of the direct sounds by virtue of their arrival times and directions provide the binaural hearing system with information about the space. LEV is a sufficiently important that it is a measure of excellence in the evaluation of concert halls. It involves reflected sound, and is therefore related to, but different from reverberation. Reverberation, a time-domain phenomenon, is present in monophonic reproduction, but envelopment is absent. LEV is only moderately represented in stereo reproduction, but through multichannel audio systems, moviegoers have been exposed to envelopment for many years, the sense of being in the space seen on the screen. Multichannel and immersive music can be very realistic when well done.
- *Immersion* is a recent addition to sound reproduction. The definition itself is vague. The currently popular concept begins with being surrounded by many loudspeakers, including overhead, each reproducing static or moving sounds. Rainstorms, fluttering leaves in a forest, jungle sounds, the sounds of a city street, applause from surrounding audience members—all "immerse" listeners. Immersion, in this sense, can happen outdoors, whereas envelopment is the impression of being within a large room. It is not envelopment, because there is no sense of a specific acoustical space created by a direct sound followed by reflections from walls, floor and ceiling. Immersion in the cognitive sense could be the classic "suspension of disbelief" when audio and visual information combine to tell a story. Other definitions extend the concept to the notion of one being "immersed" in a good story, which can be a book. Here, we will consider only the physical version of immersion: sound images perceived to originate in numerous locations.

Spaciousness is dependent on playback sound level because the illusions require that low-level reflected sounds be audible. The more of these sounds that are heard, the greater is the spatial impression. In live concert hall performances, profound spaciousness is most experienced in loud passages. In reproduction it is more apparent at high sound levels and in the absence of loud background masking noises. In car audio, spaciousness is best experienced when parked. It significantly reduces as background noise increases with speed, which is one motivation for premium systems to upmix to multichannel, and to adjust sound level as a function of speed.

4.1.1 Musical Instruments, Microphones and Rooms

Musical instruments radiate sound in all directions with many frequency-dependent directional patterns. High frequencies from violins radiate vertically away from the top plate of the instrument, while other frequencies radiate in other directional patterns. Those from a trumpet project strongly outward from the bell of the instrument. Other instruments may radiate almost omnidirectionally, or with a pattern of a dipole radiator (figure eight) over different frequency ranges. In a concert hall, one of the functions of the stage enclosure and the immediate surroundings of the orchestra is to collect these sounds and reflect some of them back towards the musicians so they can hear each other. The other function is to communicate a blended mixture of orchestral sounds into the audience through the reflected, relatively diffuse, sound field.

Beyond the first few rows in a concert hall, most of what one hears is reflected sound. Two ears and a brain, binaural hearing, in the live venue are able to separate the direct sound from later arrivals: the sound source from the room (Section 1.6). A microphone at the same location combines them, without discrimination, so such recordings, when replayed, sound far too reverberant. This is why recordings are made using different techniques, with multiple microphones closer to the instruments and voices. To record successfully in audience locations one needs binaural methods (Section 3.3).

When microphones are located close to voices and musical instruments they respond to only a small sample of the blended sounds that are delivered to the audience. Figure 4.1 shows the frequency-dependent directivity of a violin. Examining this and other musical instruments, it is evident that no single axis is an adequate representation of the timbral identity as heard by the audience in a reflective performance venue (Benade, 1985), Meyer, 1978, 1993). One of the skills of recording engineers is the choice, placement and equalization of microphones so that what we hear in recordings is a fair representation of what we hear in real life. We do not listen where microphones are located, but over time we have adapted to many of these perspectives. Lavalier microphones are universal on TV news/interview programs, often placed under the chins. My observation is that not all are well equalized, and some voices are muffled, less intelligible.

The common practice of placing microphones above the violin section has been blamed for them sounding strident in recordings—note the 2500–5000 Hz beam of sound. The recording engineer should equalize to compensate when this happens.

Different microphone techniques are needed for different instruments because the portion of the sound field sampled by the microphone often exhibits spectral biases that are not in the overall integrated sound of the instrument; think of a grand piano. Recording studios are deliberately reflective, and the best of them have quite diffusive sound fields to spatially integrate the differing off-axis sounds of musical instruments into a pleasing whole. Benade (1986) argued that recordings should provide listeners with a "room average", accomplished with what he called "reasoned miking". This, he said, would be liked by concertgoers, but often disliked by audiophiles, who have been conditioned with close-miked, highly processed, reflection-free sound.

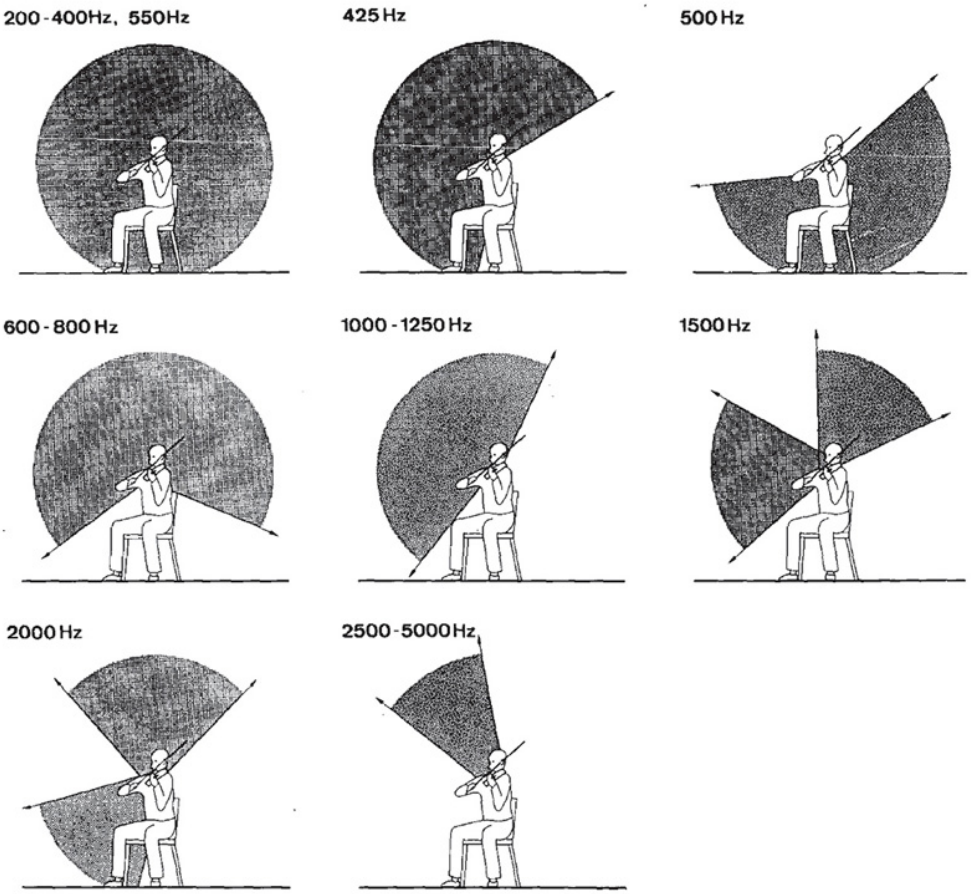


FIGURE 4.1 *Illustrations representing the sound radiated in different directions at different frequencies by a violin. It is clear that no single microphone location can capture a totally balanced spectrum. From Meyer (1993, figure 11).* ⏎

Popular music has embraced increasing numbers of "electric" instruments, which deliver their sounds to the mixing console by wires, not microphones. Guitars, basses, and keyboards are common examples. Add the content now available from sampling and synthesizers and the need for a traditional recording studio has been much reduced. The evidence is in the proliferation of home and project studios, where recordings are made and mixed in the same room. Computers are replacing consoles. The professional audio world is evolving.

4.2 THE MONOPHONIC ORIGINS

In the beginning there was monophonic, single-channel sound, and it was revolutionary. People were so impressed that sound could be captured and reproduced that they succumbed to cleverly arranged "live vs. reproduced" demonstrations, implying that something resembling perfection had been achieved over a century ago (Figure 4.2).

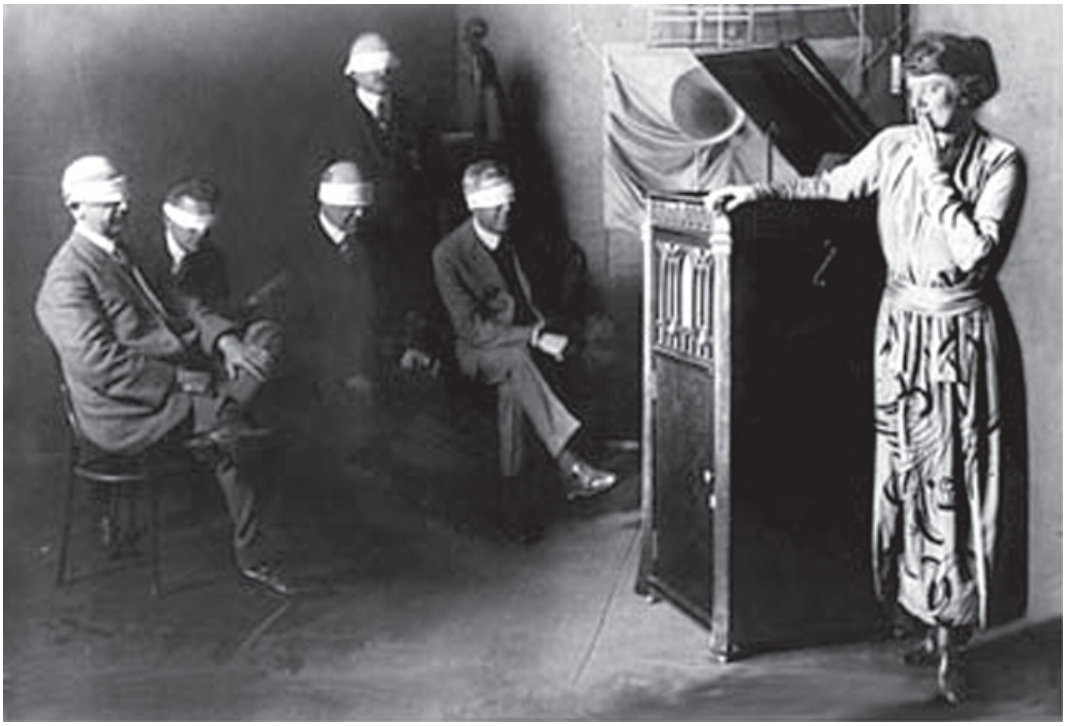


FIGURE 4.2 *Singer Frieda Hempel stages a tone test at the Edison studios in New York City, 1918. Care was taken to ensure that the test was "blind", but it is amusing to see that some of the blindfolds also cover the ears. Courtesy of Edison National Historic Site, National Park Service, U.S. Department of the Interior.* ⏎

Edison, in 1901, claimed that the phonograph had no "tone" of its own. To prove it, he created a traveling show in which his phonograph was demonstrated in "tone tests" that consisted of presentations with a live performer. Morton (2000) reports, "Edison carefully chose singers, usually

women, who could imitate the sound of their recordings and only allowed musicians to use the limited group of instruments that recorded best for demonstrations" (p. 23). Of a 1916 demonstration in Carnegie Hall before a capacity audience of "musically cultured and musically critical" listeners, the *New York Evening Mail* reported that "the ear could not tell when it was listening to the phonograph alone, and when to actual voice and reproduction together. Only the eye could discover the truth by noting when the singer's mouth was open or closed" (quoted in Harvith and Harvith, 1987, p. 12).

Singers had to be careful not to be louder than the machine, to learn to imitate the sound of the machine, and to sing without vibrato, which Edison (apparently a musically uncultured person) did not like. There were other consequences of these tests on recordings. The low sensitivity of the mechanical recording device made it necessary for the performers to crowd around the mouth of the horn coupled to the diaphragm driving the cutting stylus. Instruments that could play especially loud were favored. Because the promotional "tone tests" were of solo voices and instruments, any acoustical cues from the recording venue would reveal the recording as being different from the live performer in the demonstration room. Consequently, in addition to employing what was probably the first "close microphone" recording technique, Edison's studios were acoustically dead (Read and Welsh, 1959, p. 205). All of this is quite an elaborate deception at the very origin of the audio industry, but for listeners there had to be an element of amazement that anything approximately musical could be recorded and reproduced.

In monophonic reproduction, reflections within the listening room provide the only spatial setting for what is reproduced. Recordings contain only reverberation: a time-domain phenomenon. Some enthusiasts aimed the loudspeaker towards a wall or into a corner to add more spatial complexity, softening the single "point source" image. Loudspeakers of this period were generally quite distinctively colored, and "good tone" was the pleasing objective. It was a beginning.

4.3 STEREO ADDS A CHANNEL AND BECOMES THE STANDARD

Monophonic reproduction conveys most of the musically important dimensions: melody, harmony, timbre, tempo and reverberation, but no sense of multiple images on a soundstage or of spatial impressions. In 1930s the essential principles by which the missing directional and spatial elements could be communicated were understood, but there were technical and cost limitations to what was practical. It is humbling to read the wisdom embodied in the Blumlein-EMI patent (Blumlein, 1933) applied for in 1931 describing 2-channel stereo techniques that would wait 25 years before being popularized. It is especially interesting to learn that the motivation for stereo came in part from a desire to improve cinema sound (Alexander, 1999, pp. 60, 80). Cinema played a central role in the development of present-day incarnations of multichannel audio and now immersive audio. It is curious to read purist audiophiles belittling movie sound when all of the soundstage, space and imaging effects they so treasure would be unlocked if they were to embrace multichannel audio. The sad part of that story is that the mainstream music industry has dragged its feet at every stage, beginning with stereo, resisting efforts to add channels. The objections had mainly to do with cost to the industry and to consumers.

The benefits of more channels were recognized very early in the history of audio, notably when the Bell Telephone Laboratories investigated the reproduction of a realistic auditory perspective (Steinberg and Snow, 1934). They concluded that there were two alternative reproduction methods that would work: binaural and multichannel. Binaural (dummy head) recording and headphone reproduction is the only justification for the "we have two ears, therefore we need two channels" argument.

By 1953 ideas were more developed and, in a paper titled "Basic Principles of Stereophonic Sound", Snow (1953) describes a stereophonic system as one having two *or more* channels and loudspeakers. He says:

"The number of channels will depend upon the size of the stage and listening rooms, and the precision in localization required." He goes on:

> For a use such as rendition of music in the home, where economy is required and accurate placement of sources is not of great importance if the feeling of separation of sources is preserved, two-channel reproduction is of real importance.

So, two channels were understood to be a compromise, "good enough for the home" or words to that effect, and that is exactly what we ended up with. The choice had nothing to do with aesthetic ideals, but with technical reality that, at the time stereo was commercialized, nobody knew how to store more than two channels in the groove of an LP disc.

Vermeulen (1956) understood what stereo could and could not do, saying:

> Although stereophonic reproduction can give a sufficiently accurate imitation of an orchestra [the soundstage], it is necessary to imitate also the wall reflections of the concert hall, in order that the reproduction may be musically satisfactory. This can be done by means of several loudspeakers, distributed over the listening room, to which the signal is fed with different time-lags. The diffused character of the artificial reverberation thus obtained [the envelopment] seems to be even more important than the reverberation time.

He was right on all counts. The numerous spatial enhancers over the subsequent years, up to and including contemporary stereo-to-multichannel upmix algorithms, absolutely support his insight.

The choice of two channels was in part decided by the prevailing medium, LPs, but there is a story of its origins, well told elsewhere (https://en.wikipedia.org/wiki/Stereophonic_sound [accessed June 2014]). Clark et al. (1957) lays out an especially detailed technical perspective.

Stereo is designed for loudspeakers. There is a generally agreed upon symmetrical playback configuration of left and right loudspeakers arranged at ±30° from center for a forward-facing listener, although reduced angles are common (Figure 4.3).

To hear the stereo illusion—phantom images between the loudspeakers—the listener must be in the symmetrical "sweet spot". When listening off-center the music continues essentially unchanged, but the soundstage is distorted. A sense of ambiguous space is likely to be preserved. The recording method and style of music are determining factors. Spacious classical music is more forgiving than most popular recordings. It is amusing that in many store demonstrations there are two chairs side by side, with the axis of symmetry falling between them. An attempt at marital diplomacy fails to deliver proper stereo to either listener.

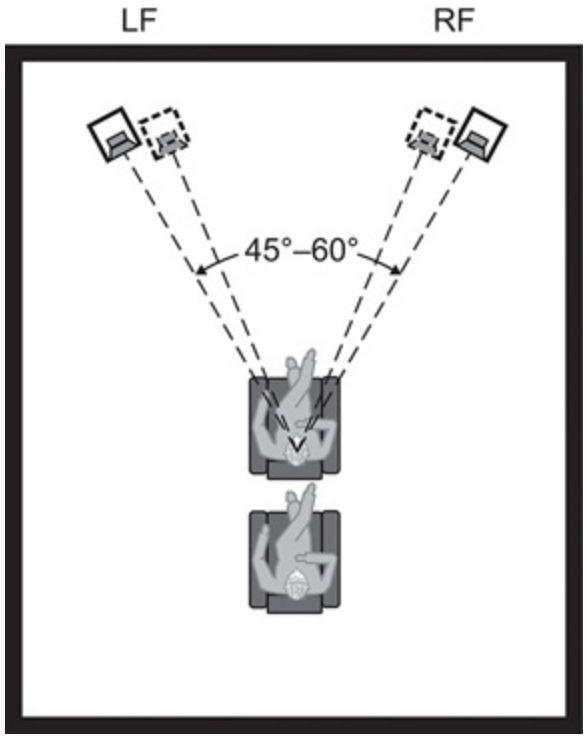


FIGURE 4.3 *Two-channel stereo. The ±30° arrangement is the prevailing standard for music recording and reproduction. Many domestic systems employ a smaller separation, especially if video displays are involved, especially TV sound bars.* ⏎

Again, live vs. reproduced comparison demonstrations of loudspeakers were used to try to persuade listeners that perfection had been achieved. One conducted by RCA in 1947 used a full symphony orchestra (Olson, 1957, p. 606). Wharfedale in the 1950s had a version (Briggs, 1958, p.

302). Acoustic Research conducted several live vs. reproduced comparisons in the late 1950s and 1960s using a string quartet and a guitar.

These were not the "party trick" events that Edison demonstrated but serious proceedings conducted by knowledgeable people in respected companies. Villchur (1964a, 1964b) provides insight into the process used by Acoustic Research. There were elaborate preparations to improve the chances of success in these demonstrations. Care was taken in the microphone placement to find a spectral signature that ensured a good reproduction, along with judicious use of equalization. The need for close microphones to avoid picking up room sound at the recording site made matters more difficult. This means that the tiny sample of the total radiated sound of the instrument would be used to represent the whole sound field.

These events avoided the circle of confusion problem (see Figure 1.8) because custom recordings were created for reproduction through specific loudspeakers at those events. But there is another factor. Most of the comparison demonstrations took place in concert halls or auditoriums. The hall sound is a huge embellishment to live performances, and it similarly benefits these reproduced sounds. Villchur (1964a) recognized this, saying: "Playback at the recording site may not sound as similar to the live music as at the actual concert, because the reverberation which will cloak both live and reproduced music equally is missing." His choice of the word "cloak" is interesting, considering the normal meaning of the verb: to hide, cover, or conceal.

In spite of the trustworthy impressions given by these exhibitions, they were, in fact, contrived marketing exercises. Not everyone was persuaded that those loudspeakers were good enough, but many people were. Loudspeakers that were criticized in properly conducted listening tests in small rooms using commercial recordings were judged to sound fine when evaluated against live instruments in public comparisons. See Figure 1.3(b) for measurements on the loudspeakers used in the Villchur demonstrations —it was seriously criticized in the author's first blind listening tests in 1965–1966. Are we now wiser, more aurally acute, and less likely to be

taken in by a good demonstration? What really matters is how commercial recordings sound in small rooms, and that is not what was demonstrated.

As for stereo itself, early recordings suffered from "ping-pong" gimmicky demonstrations, but some recordings hinted at an interesting future. All audiophiles of the time were not convinced though, and for everybody the additional cost of stereo was a factor. Dual-inventory mono and stereo versions of LPs persisted well into the 1960s, until customers learned that mono phono cartridges that allowed no vertical movement damaged stereo LPs, and replaced them.

4.3.1 The System of Stereo—How It Works

Stereo is the default format for recorded music and much more. The word "stereo" derives from the Greek word meaning "solid", as in three-dimensional. However, the reality we hear from two loudspeakers, as enjoyable as it is, is less than advertised. The special binaurally processed schemes described in Section 3.3.4 are very different.

Monophonic reproduction delivers the musical information, including reverberation, a time-domain phenomenon. The distinctive deliverables for stereo are what is commonly called "soundstage and imaging" and a sense of spaciousness. These are worthy additions. For a listener seated in the symmetrical sweet spot, there can be localizable images floating in the space between the loudspeakers, giving the impression of a band or an orchestra. There may be an impression of being in a space of some kind, a concert hall for the classics, or most often just any space larger than the listening room. These spatial effects can extend beyond the span of the loudspeakers, depending on the particular loudspeaker/room interactions and/or creative manipulations in the recording studio. It is very entertaining.

Anyone sitting away from the stereo seat cannot appreciate the recorded soundstage and imaging. For those listeners, the musical content is retained, there may be a crude sense of a left and right soundstage, and they get a spatial impression, but it is not the intended stereo illusion. Most casual listening probably falls into this category. Fortunately, music survives.

Let us now examine the mechanics of how a soundstage is constructed for a properly located listener. The first-arriving direct sound is the dominant factor for localization, and it is also fundamentally important to the perception of timbre. Sound images localized to the left or right extremes of the soundstage are essentially monophonic sounds—single loudspeakers operating alone. The center phantom image is perceived when identical sounds are radiated by both loudspeakers, which is *double-mono*. Images filling in the spaces to the left and right of center are also double-mono but with different amplitude and/or time differences between the channels being used to place the images across the soundstage.

The soundstage that is delivered can be the result of placing a microphone array in front of a group of musicians in a performance venue or by manually panning individually recorded musicians (tracks) to specific locations across the soundstage—most recordings use image panning. Synthesized reverberation and spatial effects are added to put the musicians into the desired spatial contexts.

To move an image towards one side or the other from center, the sound level of one channel can be increased and the other reduced to maintain constant overall loudness. Motion is towards the louder side. This mixing technique employs pan pots—panoramic potentiometers—to position images in the soundstage. The images tend to be of close-miked or direct-pickup sounds. About 20 dB *interchannel amplitude difference* (IAD) is required for an image to be perceived at the extreme of the soundstage—at a loudspeaker. An IAD "hard-panned" sound image is essentially monophonic and will be heard at the loudspeaker location by listeners over a large area. There will be two such well-localized images, the L and R loudspeakers.

Interchannel time differences (ITD) can also be used, in which case a delay of approximately 1.2 ms is sufficient to displace an image fully to the side. Motion is towards the earlier side. Actual values to achieve a specific location depend on the spectrum and temporal structure of the sounds. Using ITD a hard-panned image will be perceived as such only for a

listener in the sweet spot because the sound level radiated by both loudspeakers remains the same.

Most image positioning during a studio mix is done with amplitude panning, IAD. It has been reported that amplitude-panned images are more compact than time-panned images, and as mentioned, there can be two well-localized hard-panned images at the loudspeakers for multiple listeners.

Soundstages created using stereo microphone arrays obey the same interchannel amplitude and time rules as electronically-panned images. However, because sounds are captured in an acoustical setting, more of the off-axis source sound, and some of the ambiance of the performance venue is collected along with the direct sounds. These different methods are employed, for good reason, in different types of music or styles of recording. Recording a symphony orchestra in a concert hall is done with a microphone array positioned above and at the front of the orchestra, but closely positioned “spot” microphones are often used to enhance specific solo instruments or voices, or those that lack the acoustical output, like a harpsichord, or to allow the enhancement of certain sections of the orchestra. Later, in the control room, the sounds from spot microphone pickups can be panned to the appropriate location in the orchestra, delayed to synchronize timing with the rest of the orchestra, and, if necessary, augmented by synthetic reflections to match the natural acoustics of the recording venue. Done well, the trickery is not noticed.

Spending time on a streaming service that allows one to hear the same symphonic passage as recorded by different labels at different times reveals an absolutely enormous range of interpretations of auditory perspective and sound quality. Some are not very good.

The acoustical version of amplitude panning employs coincident directional microphone pairs (e.g. Blumlein pairs) to generate interchannel amplitude differences, IAD, for musicians located at different angles in front of the microphones. Spaced omnidirectional microphones generate interchannel time differences, ITD, for musicians located at different angles. Most recordings are made using spaced microphones that are also

directional, so they generate simultaneous IAD and ITD. Ensembles or large orchestras employ multi-microphone (coincident or spaced or combinations), with different directivity patterns aimed in different directions. These methods tend to be used where it is wished to include some of the room sound, and this creates challenges in finding the optimum microphone configuration and location.

Much trial and error over many years has led to some conventions of microphone arrangements that yield attractive soundstages for different kinds of band/orchestra/venue circumstances. It is part of the art of recording, and part of the expertise of recording engineers. Wittek and Theile (2002) present comparisons of several popular stereo microphone configurations, relating locations of musicians in front of the microphones to locations of the perceived stereo images. Michael Williams has long been associated with guidance for using stereo pairs to achieve specific listener experiences (e.g. Williams, 1987), which has now expanded to including three-dimensional audio (Williams, 2023). Wittek and Theile (2017) have also expanded their work in this direction. This thoughtful work was all done while listening through conventional forward-firing loudspeakers.

In summary, phantom images across the stereo soundstage are constructed from double-mono sounds from left and right loudspeakers with differing amounts of interchannel amplitude and/or time difference. The extreme left and right image locations in IAD panned recordings are substantially mono. Stereo is not an encode/decode system, it is merely a 2-channel delivery system. There are no rules, and a look at any of the numerous manuals teaching recording techniques will reveal many variations in the guidance for microphone selection, arrangement, and electronic processing to achieve a pleasing result. Nowadays, electronic processing has reached new highs in capability, and one often hears evidence of binaural processing mixed in with conventional methods. Sounds are no longer restricted to the space between the loudspeakers, but they can only be reliably heard by the listener in the sweet spot.

4.3.2 A Phantom Image Problem

Audiophiles judge the performance of stereo systems on the basis of soundstage and imaging. There are expected differences in personal preferences, some preferring "pinpoint" localizations, others not. Not often recognized is the fact that stereo soundstage images are created in a way that modifies the timbre of the images, especially that of the center image, the featured artist in most cases. Whatever else we subjectively respond to, this cannot be ignored. Stereo phantom images are not equivalent to images localized from real sources, as in the normal world. This fact creates a serious problem for the recording industry when attempts are made to combine a center channel loudspeaker with a phantom image soundstage. The real (mono center) image sounds timbrally and spatially different from the phantom (double-mono) image (King et al., 2023).

The stereo phantom center image will suffer from a significant dip in the spectrum around 2 kHz because the sound from both loudspeakers reaches both ears at different times. This interaural crosstalk cancellation is the first dip in an acoustical interference comb filter as explained in Figure 4.4.

In Figure 4.4(a) a listener receives direct sounds, one per ear, from a center channel loudspeaker. Figure 4.4(b) is the situation for a phantom center image; there are two sounds per ear, one of which is delayed because of the additional travel distance. Both of these are symmetrical situations so the sound in both ears is essentially identical. Figure 4.4(c) shows the frequency responses measured at one ear when sounds were delivered by a center loudspeaker and then by a stereo pair of loudspeakers, set up in an anechoic chamber.

Measurements were made using a KEMAR head and torso simulator (HATS) (www.grasacoustics.com/products/head-torso-simulators-kemar [accessed August 2024]), an anthropometrically and acoustically precise head and torso with ear canals terminated by microphones and appropriate acoustical impedances. The curves are *very* different. The solid curve is the "correct" one. It shows what a real center sound source delivers to the eardrum of a listener—it has the odd shape it has because of the acoustics of the external auditory system, the HRTFs. The dashed curve, for a phantom center, includes the effects of acoustical interference caused by the

acoustical crosstalk from the two loudspeakers and, the errors in HRTF processing because the incident sounds arrive at the ears from the wrong directions: ±30° rather than straight ahead. This situation is one in which there is a conflict between the timbre of the sound where it is coming from and that for the location it is perceived to come from. This is not ideal.

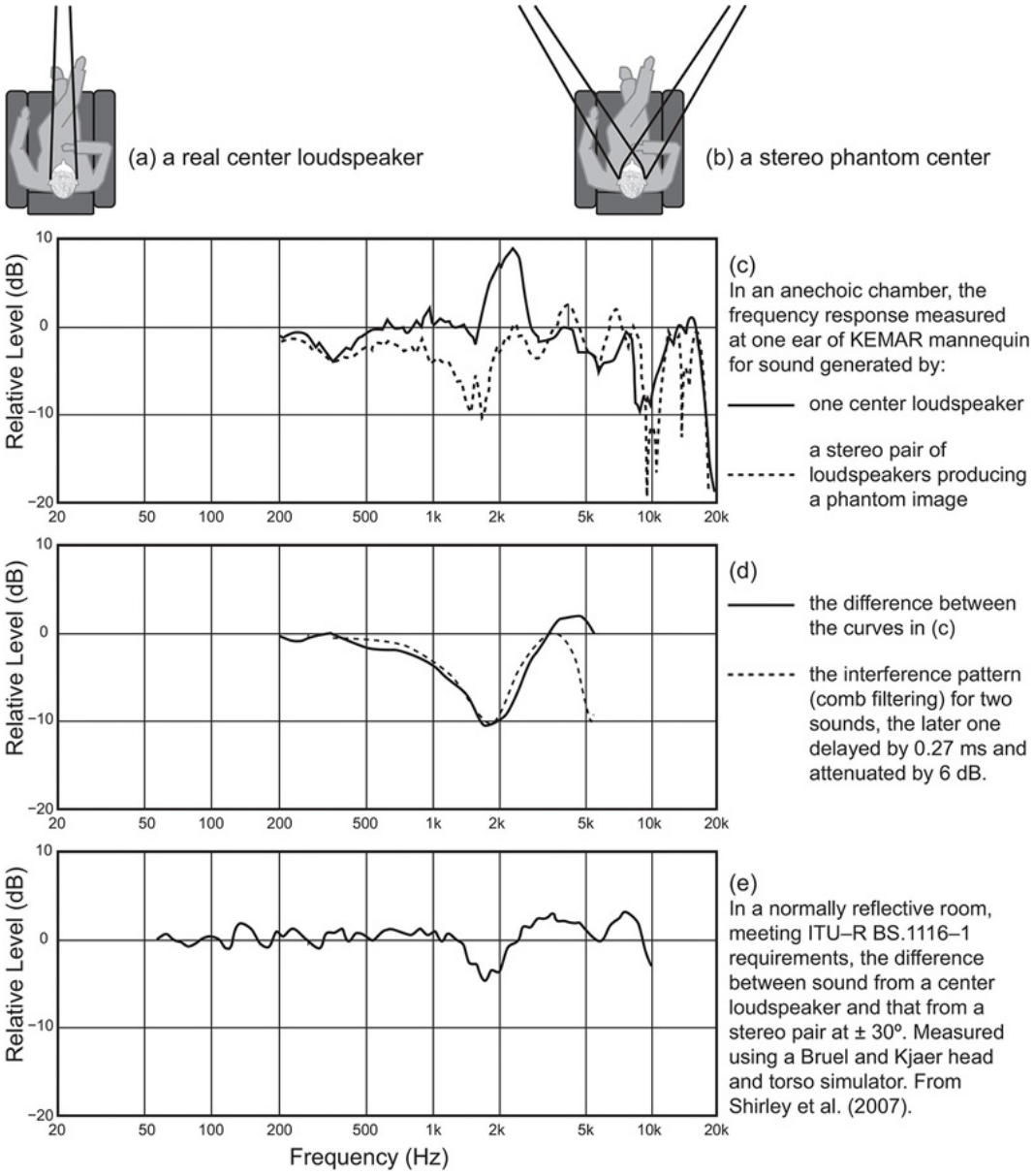


FIGURE 4.4 *Anechoic frequency-response measurements (c) made at one ear of a KEMAR head and torso simulator for sounds arriving from a real center loudspeaker, shown in (a), and from a stereo pair, shown in (b). The curves contain the axial*

frequency response of the loudspeaker used in the test, as well as the HRTFs for the relevant incident angles for that particular anthropometric mannequin. The important information, therefore, is the substantial difference between the curves: a 10 dB, octave-wide dip around 2 kHz. The smoothed difference is shown in (d). Nothing is shown above about 5 kHz because it is difficult to separate the effects of this specific acoustical interference from those of other acoustical effects. The dashed curve is the first interference dip (the first "tooth" of the comb filter) calculated for an interaural delay of 0.27 ms (appropriate for a loudspeaker at 30° left or right of center) and for an attenuation of the delayed sound of 6 dB. (e) is the same kind of measurement done in a normally reflective room, showing that early reflections within the room substantially reduce both the width and depth of the interference dip. Data from Shirley et al. (2007). ⏎

Figure 4.4(d) shows the difference between the curves, revealing the result of acoustical interference. This can be confirmed by a simple calculation. The time differential between the ears for a sound source at 30° away from the frontal axis is about 0.27 ms for an average head. A destructive acoustical interference will occur at the frequency at which this is one-half of a period: 1.85 kHz. It won't be a perfect cancellation, because of a tiny propagation loss, and a significant diffraction effect. The wavelength is just over 7 inches (178 mm) that, because it is similar in dimension to the head, will experience a substantial head-shadowing effect at the ear opposite to the sound source. There will be an interaural amplitude difference of the order of 6 dB in this frequency range. Taking a simplistic view, the dashed curve in Figure 4.4(d) shows the first cancellation dip (the first "tooth") of a comb filter with these parameters. The fit is quite good. The cause of the "dullness" in the phantom center image is destructive acoustical interference. The rest of the comb is not seen because at higher frequencies the whole situation is complicated by head

related transfer functions (HRTFs) and rapidly increasing attenuation of the delayed sound at higher frequencies caused by head shadowing.

The consequence of this is that the sound quality of the phantom center image (usually the featured artist) in stereo recordings is fundamentally compromised for a listener in the symmetrical sweet spot. Those who believe that phase shift is audible and that waveform fidelity is essential must think again, because for this sound image both are seriously corrupted. That said, for good stereo imaging the two loudspeakers must be very similar in performance. The direct, usually on-axis sounds must match, meaning that achieving identical steady-state room curves is no assurance of anything, as they are dominated by off-axis sound from the loudspeakers.

As an illustration, I recommend a simple experiment. Arrange for *monophonic* pink noise to be delivered at equal level to both loudspeakers. This test signal is available on some streaming services. When seated in the symmetrical sweet spot, this should create a well-defined center image floating midway between the loudspeakers. If it does not, something is seriously wrong. If it does, consider what you hear as you lean *very slightly* to the left and to the right of the symmetrical axis: the timbre of the noise changes. It will be more obvious the closer you sit to the loudspeakers placing you in a stronger direct sound field. This happens in reflection-controlled rooms and close (near-field) listening. In fact, it is possible to find the exact sweet spot by simply listening to when the sound is dullest. Moving even slightly left or right of the sweet spot causes the sound to get audibly brighter. It is much more exact to find the sweet spot by listening to the timbre change than by trying to judge when the center image is precisely localized in the center position. There is nothing faulty with the equipment or setup; this is stereo as it is: flawed.

Now, reproduce the pink noise through only the right or left channel —the balance control can do this. Face forward as you would normally and listen to the timbre of the noise change as you turn

slowly to face the loudspeaker and return to facing forward. The difference you hear is the error in the HRTF encoding. The correct timbre for a center image is when you face the loudspeaker. The erroneous timbre is when you face forward: for many people it is brighter, but it all depends on your individual HRTFs—your ear, head and torso shapes. But the fact that it is audibly different is a problem.

By any standards this is a huge spectral distortion, a serious fault because it affects the featured "talent" in most recordings, the person whose picture is on the album cover. Under what circumstances are we likely to hear it as shown here? Obviously, only when the direct sound from the loudspeakers is the dominant sound arriving at the listener's ears. This is the situation in many recording control rooms and custom listening rooms where special care has been taken to attenuate early reflections, as well as in all close-listening, near-field situations.

In normally reflective rooms, reflections arriving from other directions at different times will help to fill in the spectral hole because of the random acoustical interference associated with those sounds. Therefore, in normally reflective rooms this will not be as serious as the curve in Figure 4.4(d) suggests, a fact confirmed by data from Shirley et al. (2007) in the data shown in Figure 4.4(e).

The scientific literature declares that it is a clearly audible effect. Plenge (1987) explains it, as done here, and goes on to relate it to some theories of timbre perception, pointing out along the way that most listeners do not sit still and are obviously not bothered by the consequent changes in sound image location and timbre. However, dedicated audiophiles and recording engineers who gravitate to the stereo seat are exceptions. Augspurger (1990) describes how easy it is to hear the effect using 1/3-octave bands of pink noise, observing a "distinct null at 2 kHz" (p. 177). Pulkki (2001) confirmed that the comb filter was the dominant audible coloration in anechoic listening to amplitude-panned virtual images, but that it was lessened by room reflections. Listeners in experiments by Choisel and

Wickelmaier (2007) reported a reduction in brightness when a mono center loudspeaker was replaced by a stereo phantom image (their figure 4). Listeners in the Shirley et al. (2007) experiments not only heard the dip seen in Figure 4.4(e) but demonstrated that it had a significant negative effect on speech intelligibility. This is hardly the epitome of sound reproduction technology.

There is a dilemma: the 2 kHz dip exists only for the listener in the sweet spot, but the sweet spot is the only location where the soundstage is accurately perceived. If the dip is boosted by equalization, the boost is added to the sound radiated in all directions by both loudspeakers and will be a resonant coloration in the direct sound for listeners in all other locations, and the indirect sound for all listeners. This is clearly not a good idea, but it is often done.

Recording engineers doing multichannel mixes face another dilemma: the significant perceptual difference in sound and spatial qualities of a real and a phantom center image, which will be discussed in the following sections. King et al. (2023) address this in some detail.

4.3.3 Loudspeakers as Stereo Image Stabilizers

Moving away from the sweet spot, the phantom center image—the featured artist—moves with you, quickly jumping to the nearer loudspeaker. Alleviating this problem has been a longstanding challenge. For a single listener, the balance control can compensate somewhat for an off-center listening position.

In the early days of stereo Hugh Brittain, recognized for introducing metal cone loudspeakers while he was at GEC in England, promoted the idea of aiming the loudspeakers so that their axes crossed in front of the listener. His name became associated with that arrangement. Knowing that loudspeakers are somewhat directional, especially at high frequencies, the idea was that as one moved away from center the more distant loudspeaker would get louder and the nearer one would be reduced in level, thereby achieving a crude amplitude panning to compensate for the delay error (time/intensity trading); the phantom image would be perceived to move

less. Obviously, with typical cone/dome loudspeakers it only works at mid-to-high frequencies and the off-axis behavior of loudspeakers is unpredictable; sound quality is at risk. Added to this, the time/amplitude trading ratio is different for sounds of different spectral and temporal structure and images tend to break up. So, it is not a solution.

Mark Davis (1987) applied science to the problem. He developed the dbx Soundfield loudspeaker. This was an arrangement of 14 drivers on the four faces of an enclosure that, in acoustical summation, radiated a 360° sound field in which the asymmetrical directivity had been carefully tailored. The purpose was to deliver direct sounds to off-center listeners that stabilized the location of the phantom center image. It was based on empirical data relating delay and amplitude panning, and differed from other attempts in that it was effective over a wide bandwidth. In addition to this direct sound feature, the loudspeaker radiated generously in all directions, delivering what many listeners find to be a pleasant spaciousness. My 1986 exposure to the product was limited to the Soundfield 10 model in which the transducer count was reduced to 8. In any array more transducers are better, but cost and complexity are limiting factors.

Unfortunately, the spatially averaged computer data in my file has been lost, leaving only a single listener-axis frequency response that is not very smooth because of acoustical interference from the sparse transducer array. Polar plots confirmed Davis's published goals. Listener notes from blind tests indicated that it generated an attractively spacious illusion, with soft and reasonably stable phantom images. Cost was a problem, so even simpler implementations followed, and then the product disappeared.

It is intriguing to contemplate having an array loudspeaker capable of broadband image stabilization without having the accompanying multidirectional radiation: a fully steerable array. Today's technology using many drivers, each independently driven and signal processed, would be capable of radiating almost any directional pattern controlled by a handheld wireless device. Products of this kind are entirely feasible, and if offered in the marketplace, very high prices can be anticipated, which is likely why

they do not exist. Sitting in the stereo seat can be a significant cost-saving decision.

Meanwhile, countless loudspeaker designs have attempted to present a more gratifying sense of soundstage, imaging, spaciousness, and envelopment—whatever individual listeners prioritize. What does one say about a system that accommodates loudspeakers having directional characteristics ranging from omnidirectional, through bidirectional in-phase (bipole), bidirectional out-of-phase (dipole), predominantly backward firing, and predominantly forward firing with different dispersion patterns? The nature of the direct and reflected sounds arriving at the listeners' ears from these different designs runs the entire gamut of possibilities, and yet one finds intelligent advocates for them all. Given that the recording itself is the dominant influence in imaging and soundstage presentation, one must conclude that stereo is not a system, but a basis for experimentation. It stimulates endless discussions in audio forums.

4.4 EVALUATING SOUND QUALITY AND IMAGING IN STEREO AND MONO

The importance of loudspeaker directivity and the associated early reflections in listening rooms is not disputed. Reviews and audio forum discussions reveal high levels of interest in the topic, but scientific investigations of the phenomenon are scarce. Doing it well is a lot of work, but it seemed to be worth the effort. Figure 4.5 illustrates the three basic categories of loudspeaker directivity and the influence they have on the strength of adjacent wall (and other) reflections. From what we know of binaural hearing, it is logical to think that these strong first-order lateral reflections must influence sound quality, soundstage and imaging in both positive and negative ways. The following is a serious look at what the effects are; sometimes intuition is wrong.

At the time the following experiments were arranged (Toole, 1985, 1986), there were no omnidirectional loudspeakers of competitive sound quality that were available to me, so two versions of forward-firing

loudspeakers were selected: a simple two-way of popular configuration, and a three-way, offering two distinctive off-axis behaviors. The highly regarded Quad ESL 63 represented the dipole category, even though it was not a perfect dipole because it attenuated the rear radiation at middle and high frequencies.

The tests were done many years ago, but at that time stereo was in its prime, and the selected listeners were very experienced and knowledgeable. The questionnaire was intended to elicit as much information as possible from the listeners and much was revealed. After 40 years, the results still generate animated discussions, including flak, in audio forums.

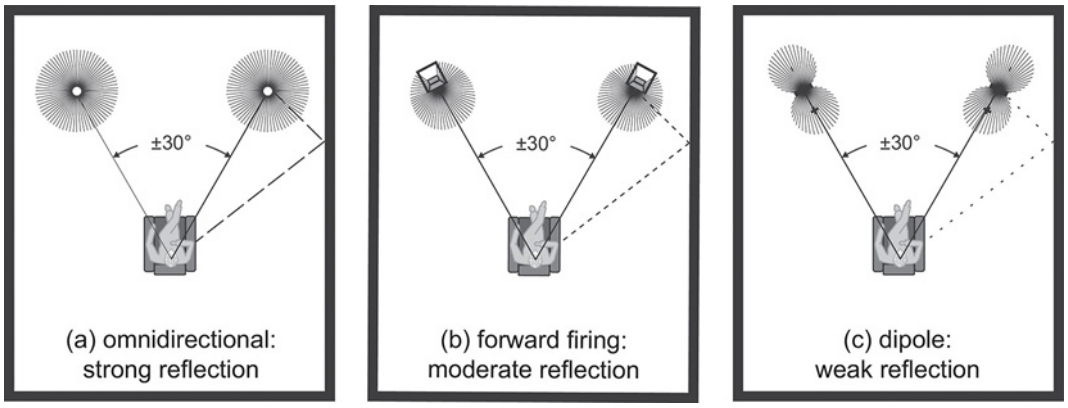


FIGURE 4.5 *Floor plans showing mid-frequency sound radiation patterns from idealized (a) omnidirectional, (b) forward-firing and (c) dipole loudspeakers. Also shown are indications of strong side wall reflections from the omnidirectional loudspeaker, moderate reflections from the forward-firing loudspeaker and weak reflections from the dipole, because of the orientation of the null in its radiation pattern. A perfectly flat reflecting wall is assumed.* ⏎

Figure 4.6 shows the room layout for stereo comparisons of two cone-and-dome forward-firing loudspeakers and a full-range electrostatic dipole loudspeaker. The room was the prototype IEC 268–13 (1985) recommended listening room. The program was muted while the turntables were manually rotated into position. The presentations were randomized. The photograph

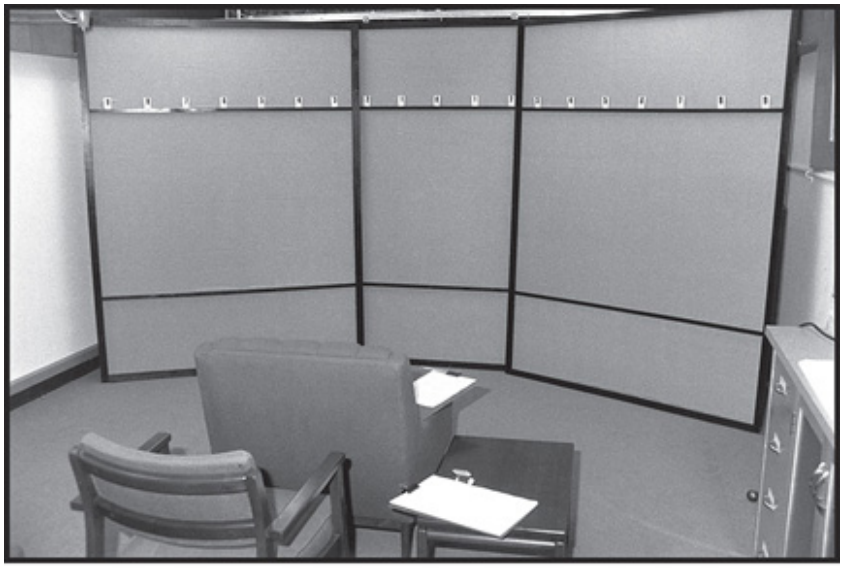
of the room shows the acoustically transparent screens, with reference markings to help listeners compare image locations and soundstage dimensions, and the two chairs, with the rear one higher than the front one to reduce shadowing effects.

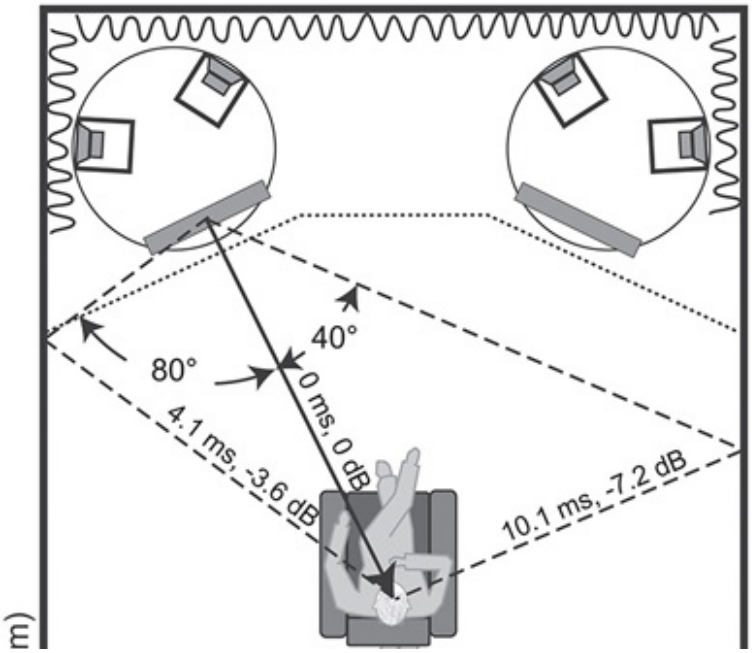
All three loudspeakers in Figure 4.7 had reasonably flat, smooth, on-axis (direct-sound) performance. The right half of the figure shows the author's crude estimate of the sounds arriving at the ears of the listener in the front seat. Unfortunately, 40 years after the fact, there is no data on the sounds radiated at the specific off-axis angles shown in Figure 4.6, so the 60°–75° data were used as a substitute for the required 80° off-axis adjacent-wall curve, and the 30°–45° data were used to represent the 40° opposite-wall curve. The curves for the specific angles, 40° and 80°, would have been lower than those shown but the comparative levels are probably realistic. These curves were lowered by the amount required by the inverse-square law, 3.6 dB for the adjacent-wall reflection and 7.2 dB for the opposite-wall reflection—both compared to the direct-sound arrival, the on-axis curve. The curves are truncated below 500 Hz because at lower frequencies, room modes and adjacent-boundary effects progressively dominate.

- The Rega Model 3 was a two-way design, 8-inch (200 mm) woofer and 1-inch (25 mm) tweeter. It shows increasing directivity of the large woofer up to the crossover to the tweeter around 3 kHz. The uneven off-axis performance is typical of many such loudspeakers.
- The KEF 105.2 was a three-way design: 12-inch (300 mm) woofer, 5-inch (110 mm) midrange, and a 2-inch (50 mm) tweeter. This loudspeaker had wider mid-frequency dispersion than the two-way but also exhibited uneven off-axis performance.
- The Quad ESL 63 was a full-range electrostatic modified dipole, employing a diaphragm subdivided into areas driven in a manner to approximate a spherically expanding wavefront. The center circle, the "tweeter", was about 3 inches (76 mm) in diameter. This loudspeaker was the most directional in the forward direction, with relatively

uniform 30° to 45° off-axis performance but with sharp irregularities 60° to 75° off axis, even after spatial averaging.

In Figure 4.7 the higher broadband directivity of the ESL 63 results in lower-level first-order lateral reflections than either of the cone and dome designs. All three systems show gently uneven frequency responses in the estimated opposite-wall reflections, but all three are much worse in the farther off-axis adjacent-wall reflections, and the actual off-axis angle is greater than those used to create the estimates.





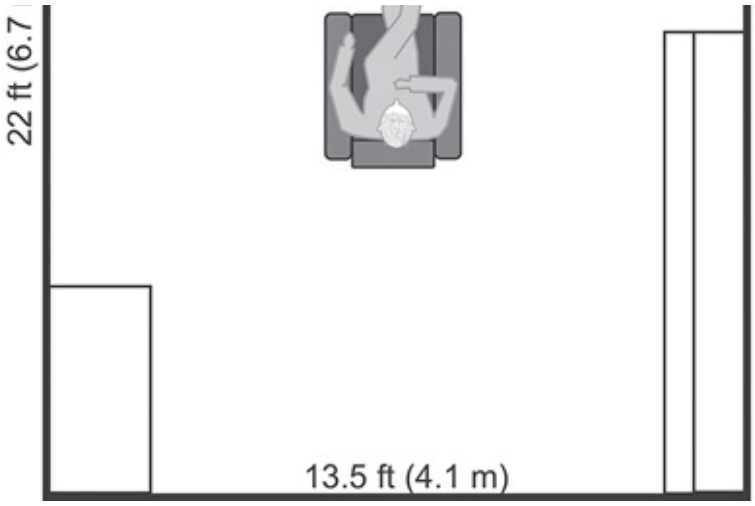


FIGURE 4.6 *The physical arrangement for the listening tests, showing the turntables used to rotate the three loudspeaker pairs into identical positions for listening. An acoustically transparent screen prevented listeners from seeing the loudspeakers. The geometry was such that the adjacent side wall reflections occur at very large angles off axis (80° and 68° for front- and rear-row listeners, respectively), and those from the opposite wall are traceable to moderate angles (40° and 31°). Medium-weight drapes covered the walls behind the loudspeakers. This would have the greatest influence on the bidirectional dipole loudspeaker, the Quad ESL 63, but the product was already equipped with absorbing pads in the rear half of the enclosure to attenuate the output above about 500 Hz. The curtains would further attenuate the rear radiation and absorb mid- to high-frequency sound escaping to the rear from the forward-firing designs. The side walls between the loudspeakers and the listeners were hard, flat broadband reflectors. The tests were done in stereo and mono, the latter using only the left loudspeaker.* ⏎

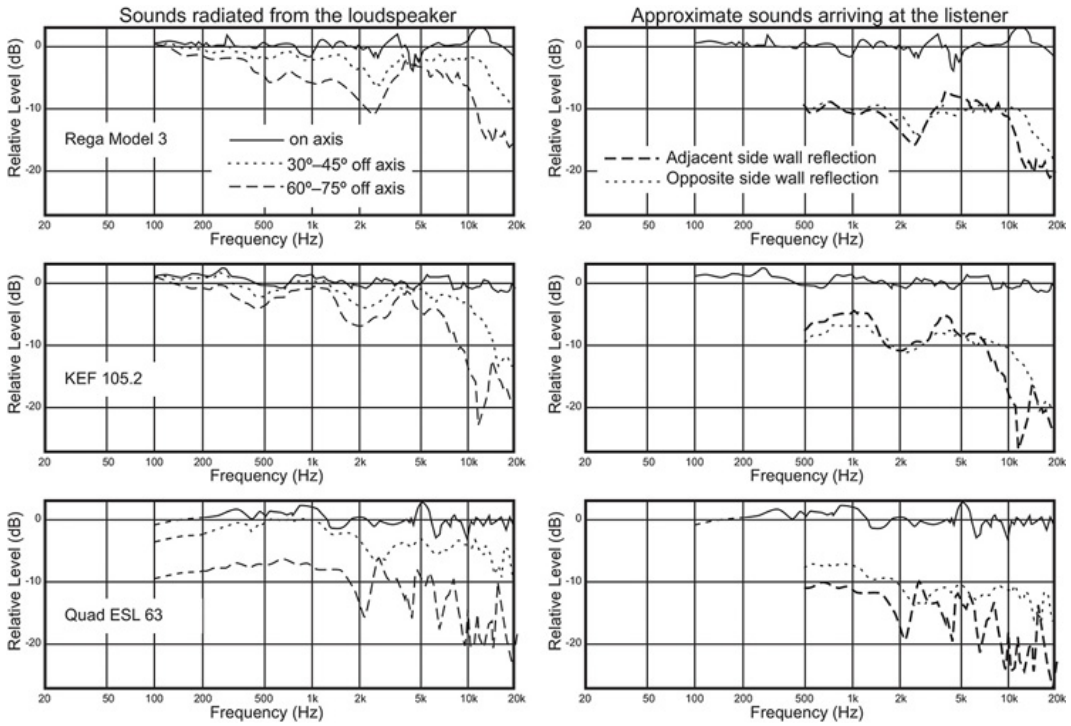


FIGURE 4.7 *The left half of this figure shows anechoic chamber measurements on the three loudspeakers (from Toole, 1986, Figure 24) with the on-axis, 30° to 45° off-axis, and 60° to 75° off-axis spatially averaged frequency responses shown.* ⏎

Ten listeners, audiophiles, audio journalists and recording professionals participated in the evaluations. These people were selected because they were knowledgeable, experienced and opinionated about audio—prerequisites for such a demanding test. They listened alone or in pairs, being cautioned to avoid verbal or non-verbal communication during the tests, and discussion between the randomized tests, which went on for several days. Generous rest periods were built into the schedule. While listening they completed a questionnaire interrogating their views on numerous aspects of sound quality and spatial qualities, and at the end requiring a "gestalt" opinion, a number on a 10-point scale that reflects their overall opinion of fidelity (sound quality), pleasantness and spatial quality. The music consisted of excerpts of choral, chamber, jazz and popular programs recorded with special care by tonmeister/recording engineering staff and students at McGill University. All were one generation from

master-tape recordings done especially for this purpose, employing known microphones and documented processing, which was minimal. The questionnaire is shown in Figure 4.8.

During the presentations, listeners responded to each of several "perceptual dimensions" in sound quality and spatial quality. At the end they were required to summarize their impressions with single number overall ratings, on a 10-point scale, for spatial quality, pleasantness and fidelity (sound quality). Comments were encouraged. Loudness levels were carefully matched, and the music selection and product presentation sequences were randomized. The experiment was done in stereo and also in monophonic form, using only the left loudspeaker. Half of the listeners started with the monophonic test and half with the stereo test. Loudspeakers were identified only as "product number", which was changed from time to time to minimize habituation (Toole, 1985, figure 2).

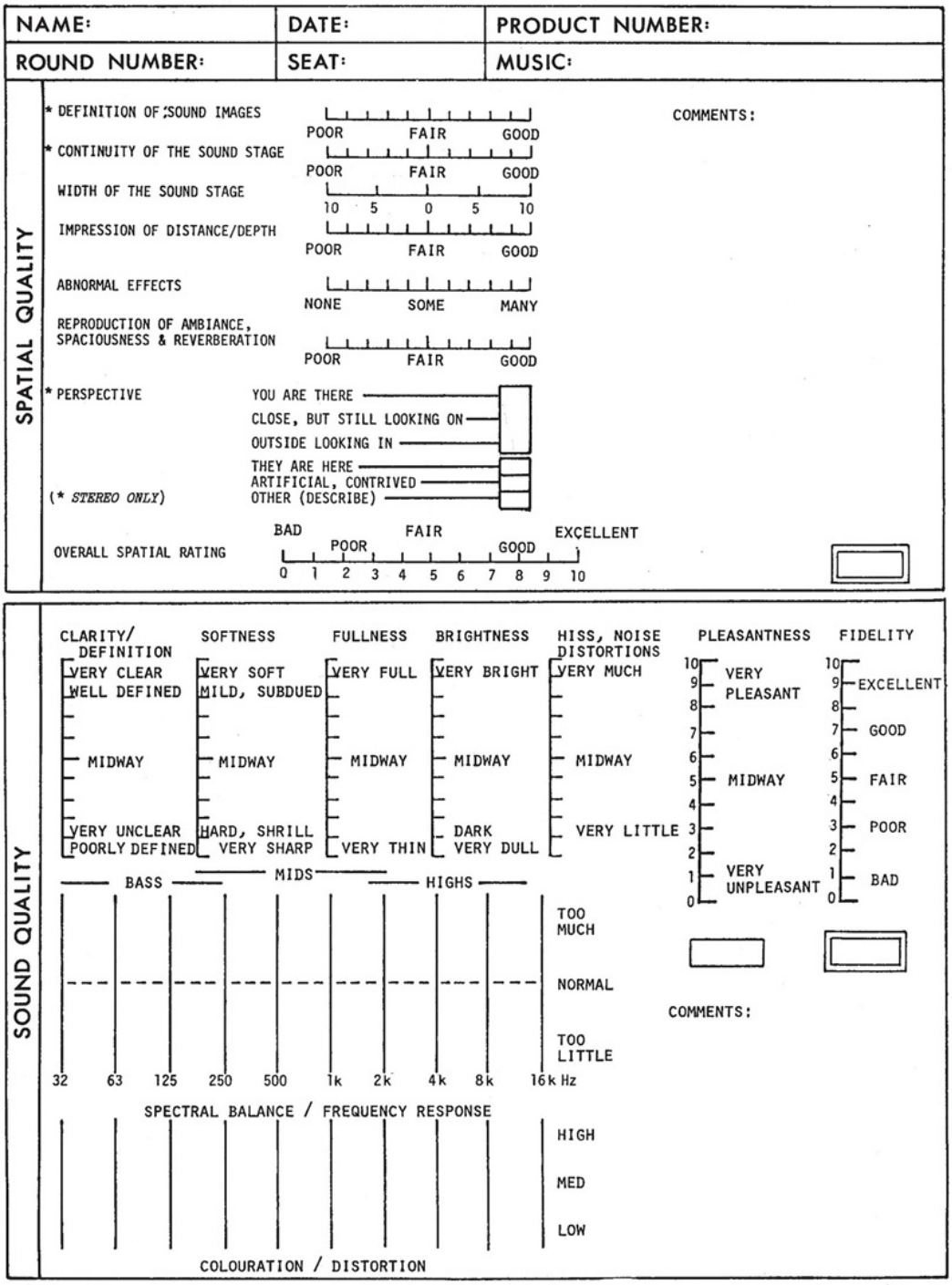


FIGURE 4.8 *The listener questionnaire used in the tests.*

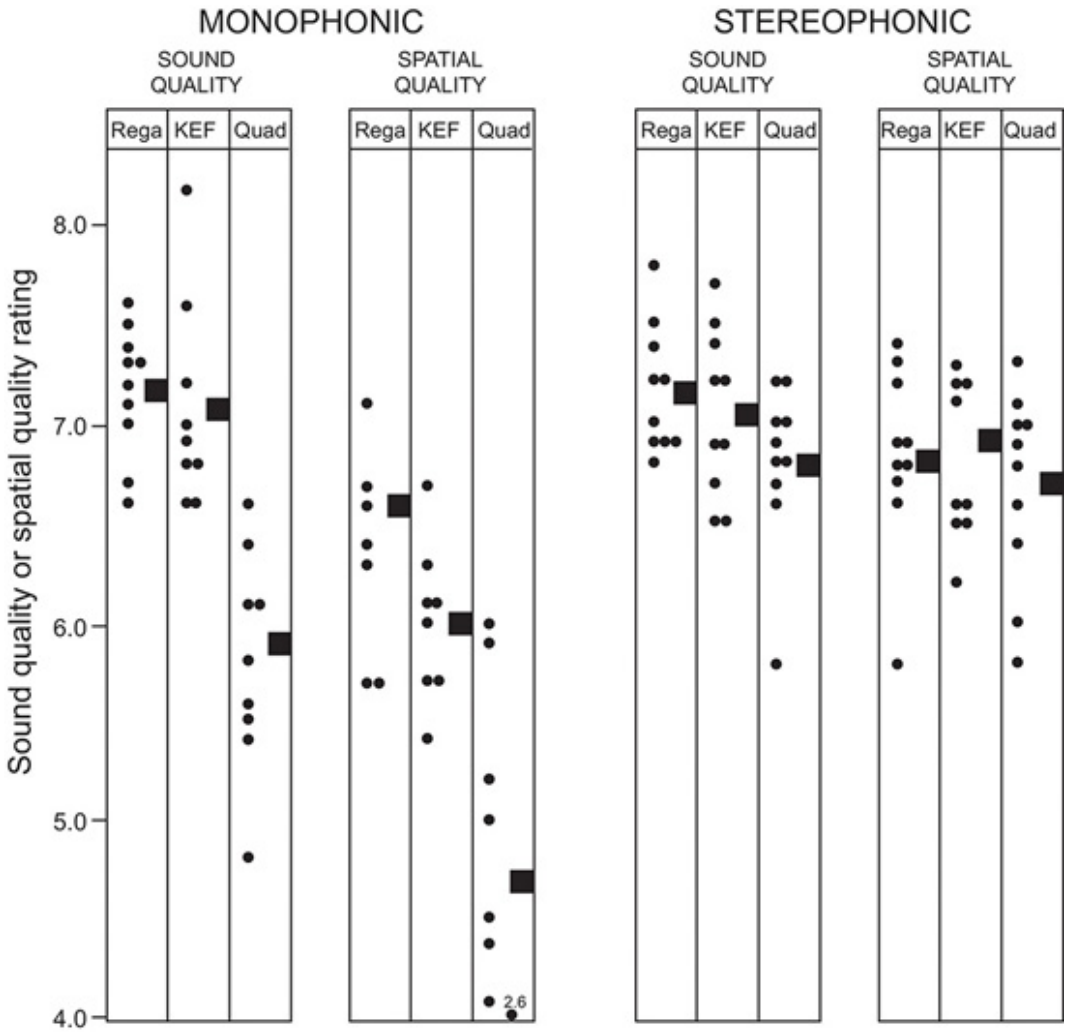


FIGURE 4.9 *Overall sound quality and overall spatial quality ratings for the three loudspeakers when auditioned in mono and in stereo. Each dot is the average of several ratings by a single listener. The ratings are averaged across the four music selections: choral, chamber, jazz and popular. On the vertical scale, 10 represented the best imaginable sound and 0 the worst. The large square symbols indicate the average of all responses for individual loudspeakers. From Toole (1985, figure 20).* ⏎

In the results shown in Figure 4.9, the first surprise was that single loudspeakers elicited strong opinions about spatial quality. Those of us who had participated in many single-loudspeaker comparisons were aware that there were differences in the perception of the spatial extent and distance of

the single sound source. To us, the most neutral-sounding loudspeakers tended to not draw attention to themselves; they almost disappeared, leaving the sense of distance to be conveyed by the recordings. The least neutral loudspeakers were localizable sources; they "sounded like" loudspeakers. In stereo this was especially noticed in recordings with hard-panned left and right images, which are monophonic signals. However, it was still a surprise that this was an impression shared by listeners not accustomed to this form of critical analysis.

In these results, spatial quality and sound quality ratings were obviously not completely independent—both exhibited similar trends. This was not expected. Was it possible that listeners could not separate them even though, consciously, most were confident that they could (the author included)?

In monophonic tests listeners reported large differences in both sound quality and spatial quality. However, in stereo listening most of the differences disappeared in the average ratings for all programs. The two highly rated loudspeakers kept their high sound quality ratings, but the loudspeaker with low ratings in mono became competitive in stereo. This was a puzzle, because it had been assumed that it was stereo spatial ratings would reveal the relative merits in terms of imaging and space.

The notion that early lateral reflections in the listening room would be audible problems was not reinforced. The reflections in the arrangement in Figure 4.6 arrived at 4.1 ms and 10.1 ms, within the range thought to be problematic; however, as shown in Figure 4.7, due to loudspeaker directivity, the reflections were attenuated by roughly 10 dB when they reached the listener. These were flat, plaster-on-block side walls—perfect specular reflectors. Yet what they contributed to the listening experience could be interpreted as being more positive than negative.

Comb filtering is apparent in measurements when direct and delayed sounds acoustically interfere with each other, so technically there is a problem. However, in perceptions by two ears and a brain, delayed sounds arriving from different horizontal angles than the direct sound are generally perceived as contributing to spatial effects—see Figure 1.5. Reflections

deliver information about the room, confirming what the eyes see, and also do more. Figure 4.4 shows that for the phantom center image, stereo introduces a problem: a massive 2 kHz dip in the direct sound. Reflected sounds in the room reduce the dip: Figure 4.4(e), improving not only sound quality but speech intelligibility. This is a parallel with the seat-dip effect in concert halls, where reflected sounds reduce the audibility of the measured direct-sound effect.

What could cause the sound and spatial quality ratings of a loudspeaker to improve when it is heard in stereo? Spatial factors were strongly in evidence in both mono and stereo tests and obviously stereo recordings add enormously to spatial and directional perceptions. Were these added perceptions masking defects in the lower-rated loudspeakers, making them less audible, or simply being a distraction?

Before addressing that question further, let us examine the raw data. Figure 4.10 shows histograms of judgments in the various categories of spatial quality listed in the questionnaire, Figure 4.8.

The soundstage and imaging ratings show a lot of similarity among the loudspeakers; it is difficult to see a clear winner or loser. In the category "abnormal spatial effects" it was interesting that the Quad was criticized for creating the impression of sounds being inappropriately close to the listener, and occasionally inside the head. These are perceptions normally associated with a dominant direct sound field, i.e. insufficient reflections to establish a sense of distance—see the following text box. Moulton (1995) noted that "speakers with narrow high-frequency dispersion … tend to project the phantom at or in front of the lateral speaker plane." This is a "depth" dimension, but with images moving closer to the listener rather than farther away. Looking at the "depth of soundstage" histogram, one could be persuaded that the cone and dome loudspeakers were more successful, but more work would need to be done to prove it conclusively.

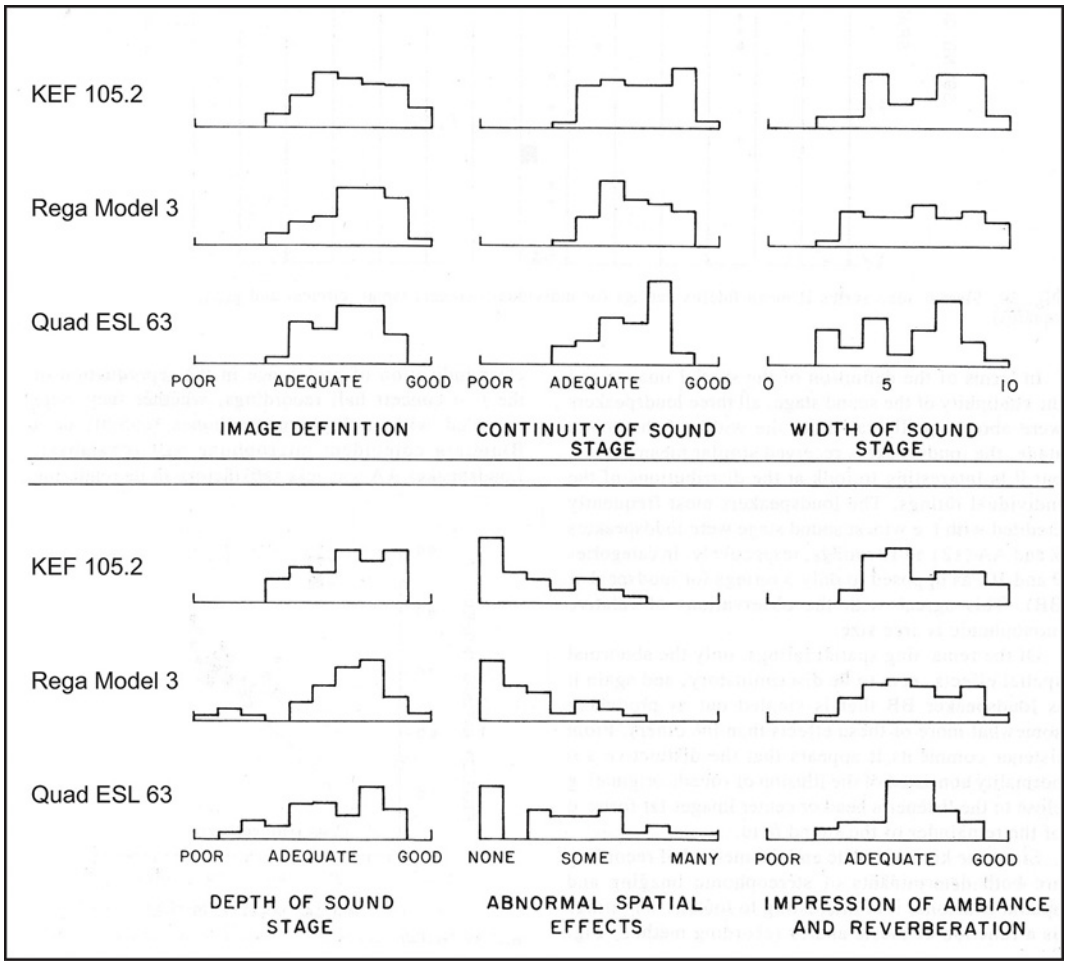


FIGURE 4.10 *Distributions of the analytical judgments in listener responses for all music selections combined. Adapted from Toole (1985, figure 22).* ⏎

In the overall description of listening perspective, for the music recorded with a natural perspective (choral, chamber and jazz), the modal listener response was "you are there" for the Rega and the KEF and "close, but still looking on" for the Quad. According to the definitions of those phrases … the Rega and the KEF gave listeners some impression of being enveloped by the ambient sound of the recording environment, with the Quad tending to separate them from the "performance" (Toole, 1986, p. 342). This all sounds very much like the influence of lateral reflections, apparent source width (ASW), early spatial impression and the associated variations in interaural cross correlation (IACC) which correlates with a sense of

envelopment. For listeners who appreciate such illusions, wide dispersion loudspeakers and some amount of room reflections appear to be desirable.

In-head localization is logical opposite of an enveloping, external and spacious, auditory illusion. Perceptions of sounds originating inside the head, which routinely occur in stereo (not necessarily binaural) headphone listening, can also occur in loudspeaker listening when the direct sound is not supported by the right amount and kind of reflected sound for the perceptual system to make a proper determination—a default localization. The author and his colleagues have experienced the phenomenon many times when listening to stereo recordings in an anechoic chamber, usually with acoustically "dry" sounds hard-panned to center or, less often, to the sides. It prompted an investigation, Toole (1970), the conclusion of which was that there is a continuum of localization experience from external at a distance through to totally within the head. It is often noted with higher frequencies, and it can happen in a normal room with loudspeakers having high directivity, or in any situation where a strong direct sound is heard without the expected reflections. In an anechoic chamber it can occur when listening to a single loudspeaker, especially on the frontal axis, in which case front-back reversals are also frequent occurrences. This phenomenon is so strong that it need not be a "blind" situation. It is very disconcerting to look at a loudspeaker and to hear the sound inside the head or behind.

These data, and the ratings of Figure 4.9, combined the ratings for all four musical selections. Perhaps something more can be learned by separately examining the ratings for each of the musical selections (and the recording techniques embedded therein). See Figure 4.11.

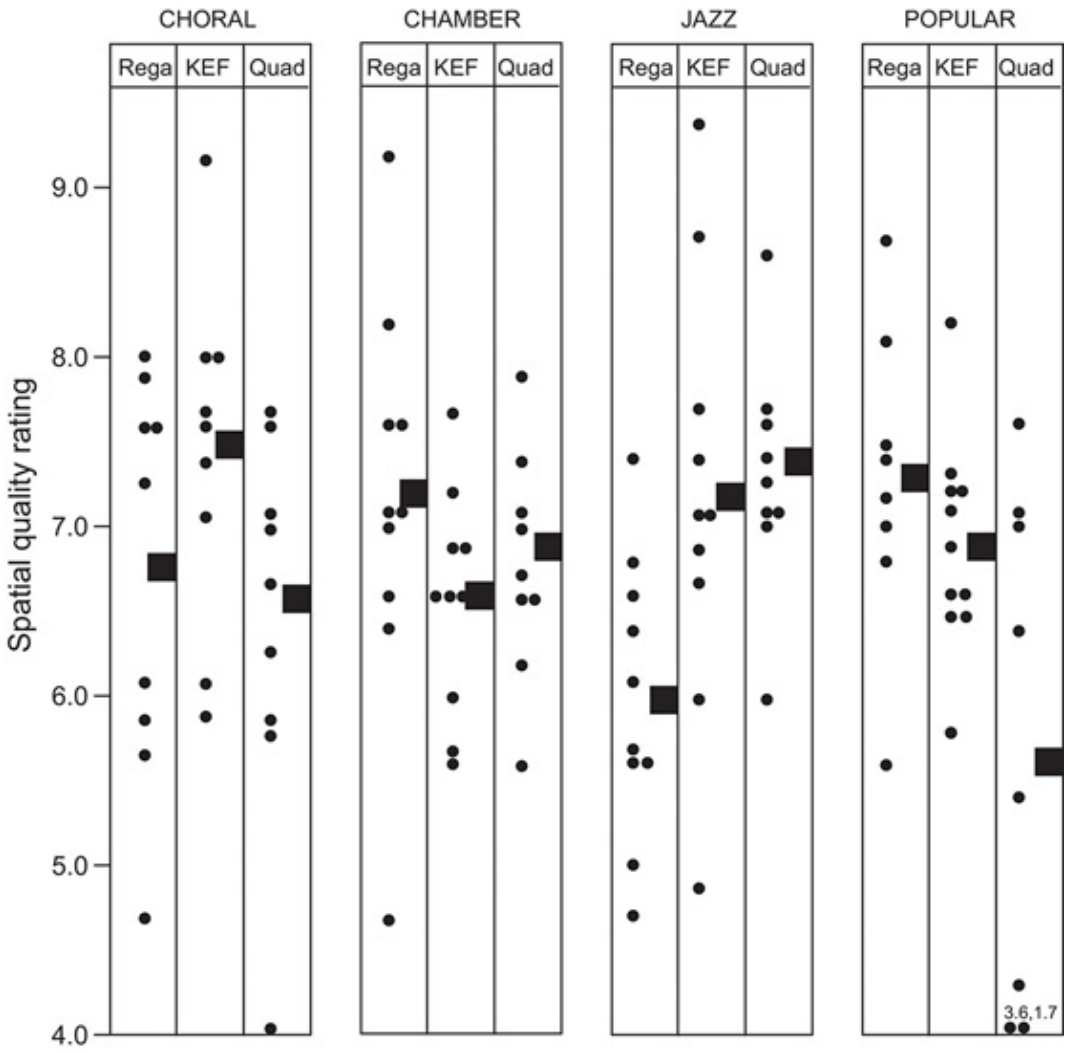


FIGURE 4.11 *Overall spatial ratings of loudspeakers in stereo, for each of the musical selections. From Toole (1985, figure 23).*

Figure 4.11 shows the spatial quality ratings for each of the music selections. *They are all different*, but there is a lot of judgment variation that reduces confidence in some of the differences. The choral selection used multiple microphones and the chamber piece employed a Blumlein coincident pair. The listening room reflections are likely to be heavily diluted or masked by the spatial information incorporated into the stereo classical recordings. Nevertheless, there are hints of preferential biases in the distributions, but they are different for the two recordings. The jazz

selection revealed lower ratings for the Rega. No explanation for this was found, but there is clearly an interaction with aspects of this specific multi-microphone recording.

In distinct contrast, the pop music selection yielded ratings for the Quad remarkably similar to those seen in monophonic listening (Figure 4.8). Why? Possibly because of the simple soundstage of the pop recording, consisting of significant amounts of hard-panned (i.e. monophonic) sound emerging from the left and right loudspeakers and a "double-mono" phantom center image. The reflected sound accompaniment for these monophonic components of the stereo soundstage is mainly the listening room. The wide-dispersion loudspeakers and the accompanying early reflections from the hard reflecting walls can be presumed to be advantageous in two respects: "softening" the localization of hard-panned left and right images and reducing the 2 kHz acoustical crosstalk interference dip in the center image—the featured artist—see Figure 4.4(e).

The principal conclusion from these results is that recording technique is probably the prime determinant of directional and spatial impressions perceived in stereo reproduction. Moulton et al. (1986) had similar experiences. In this limited evaluation it is not clear that any of these loudspeakers has a unique advantage with respect to soundstage and imaging when all recording results are combined. The directivity of the loudspeakers is obviously a factor in some recordings, as would be the reflectivity of the surfaces involved in the first lateral reflections.

Finally, there is the strong evidence that listening in mono is the most demanding test for sound quality, and, surprisingly, there is a significant spatial component in monophonic listening. Sound and spatial quality ratings looked at overall correlated with a 0.7 correlation coefficient—they track each other.

In addition to the sound quality ratings themselves, there is another relevant factor: the consistency of the ratings in the two listening modes. Toole (1985) observed that hearing loss increased the variability of subjective ratings, and that it does it differently in stereo and mono. Figure 4.12 shows that listeners with normal hearing can make good judgments in

either mode, but as broadband hearing loss increases, mean standard deviation (variability) in ratings also increases, and it increases more rapidly in stereo than in mono listening. The magnitude of hearing loss to create a noticeable degradation is quite small, not enough to alarm an audiologist. Hearing thresholds below 1 kHz are shown, but those listeners also had hearing loss at higher frequencies. High-frequency hearing loss, by itself, was not a reliable indicator in these tests, although there can be no doubt that it affects what is heard. Figure 10.7(b) shows that average audio professionals at age 40 and above may exhibit this degradation in listening ability. Bech (1989) confirmed the trend, reduced in magnitude, for a group of younger listeners.

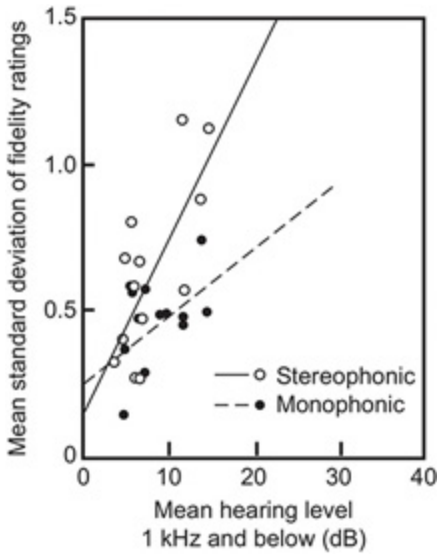


FIGURE 4.12 *A comparison of judgment variability as a function of hearing level for tests done in mono and stereo. Linear regression lines are shown. From Toole (1985).* ⏎

There can be no doubt that hearing loss (as assessed by audiometric hearing threshold determinations) and other hearing defects that are not so easily diagnosed or measured influence the formation of opinions about sound and spatial qualities. Chapter 16 will elaborate on the complexities of hearing problems. For now, a reasonable basic requirement for listeners

would be audiometric thresholds not more than 15 to 20 dB from audiometric zero (e.g. ISO 389, 2017).

In movies, the center channel is a monophonic source carrying almost all the dialog and all on-screen sounds much of the time. It and the front left and right channel loudspeakers must therefore perform well in isolation. In "middle-of-the-band" multichannel music recordings, all loudspeakers must do well as "solo" performers. Wide dispersion is of no value if the reflections cannot be heard, so this implies that some amount of room reflectivity should be present.

Subsequent stereo vs. mono tests in the intervening 40 years have not changed these conclusions. Evans et al. (2009) reviewed several studies examining the effects of loudspeaker directivity, concluding that there were several fundamental questions remaining to be properly investigated. They did note that the author's work described here "appears to be the most relevant to date, with regard to loudspeaker directivity effects."

4.5 ADDING MORE CHANNELS: A MULTICHANNEL EFFECT

The topic of multichannel systems will be discussed in an upcoming chapter, but it is interesting to look at the specific effect multichannel sound reproduction has on the perception of loudspeaker sound quality.

An elaborate comparative listening and measurement experiment was conducted by Olive et al. (2008). It had a specific purpose which will be discussed later, but the final results are highly relevant now. Figure 4.13 shows the data with the identities of the loudspeakers disguised as "Good, Better, Best" based on inspection of spinorama data and double-blind subjective evaluations.

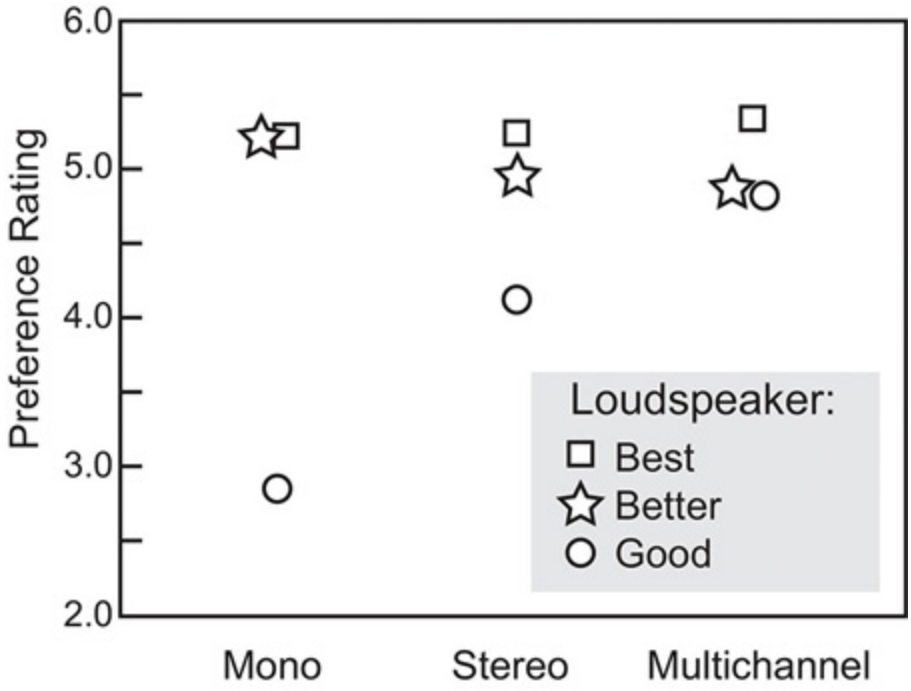


FIGURE 4.13 *The subjective ratings of a loudspeaker in three different conditions, resulting in clearly different ratings of sound quality. Comparisons were made in mono, stereo and 5-channel multichannel mode. Olive et al. (2008).* ⏎

Figure 4.13 shows a comparison of subjective sound quality ratings for a flawed loudspeaker that was improved by equalization in two different ways and then auditioned in mono, stereo and 5-channel multichannel; Olive et al. (2008). It can be seen that the overall ratings went up as performance of the loudspeaker itself was improved, but the ability to differentiate among the three conditions diminished as the channel count increased. Only in mono mode were the rating differences statistically significant.

Why is this happening? In real life, individual voices are small, monophonic sound sources, and we are well adapted to separating such sounds from those of rooms. Does introducing the sounds of recorded spaces into the listening experience cause listeners to pay less attention to sound quality? Do the room sounds in recordings that are not coherently related to those in the listening room confuse the situation beyond the

analytical ability of our hearing system? When all channels are simultaneously active these possibilities make sense, but that is not always the case.

Music and movie programs have intervals of monophonic, and nearly monophonic, sound. The talking heads on TV are a common example, and most of the dialog and much of the on-screen sound in movies is delivered by the center channel. All sounds hard-panned to a single loudspeaker are mono signals, even in contemporary "immersive" sound systems. It is therefore rational to perform the most critical loudspeaker evaluations in mono, as it is a fail-safe method from the perspective of sound quality while listening in the real world.

Nevertheless, it is interesting to contemplate how our binaural hearing systems deal with superimposed "spaces". An interesting recent study indicated that when listening to a recorded space in a real space—the listening room—it is the larger of the two spaces that is most often perceived (Hughes et al., 2016). So, the objective in sound reproduction would appear to be to employ a playback space that sounds "smaller" than any reproduced space in program material; in domestic rooms, home theaters and recording control rooms, that would appear to be the case. In contrast, large cinemas have problems reproducing the impressions of small spaces or communicating intimacy due to inherent large-room spatial cues.

Research is revealing answers and at the same time raising questions. That is the scientific method. Meanwhile audio enthusiasts pursue their hobby, often unaware of the published knowledge. Internet audio forums can be informative and fun, but occasionally frustrating. The results of NRCC and Harman monophonic loudspeaker evaluations have had their audio forum critics, who argued: why evaluate in mono when we listen in stereo? Critics assert that the imaging and soundstage aspects contributed by stereo are very demanding of loudspeaker performance and might reveal technical defects that are not revealed in monophonic evaluations.

When put to the test—proper double-blind tests as discussed above—the added perceptual dimensions in stereo turn out to be influenced as much or more by recordings than by the loudspeakers. Instead of making

loudspeaker sound quality requirements more demanding, it turns out that the recorded spatial information in multiple channels makes listeners more forgiving of loudspeaker flaws. Fortunately, well-designed loudspeakers are appreciated in all playback formats, because most programs have monophonic sound components. As discussed earlier in this chapter, stereo adds its own problems. Music is still very much enjoyed, but compromises and adaptation are involved.

Finally, advice to loudspeaker users: evaluate loudspeakers in single loudspeaker monophonic comparisons (to find out what you have). Then, if they are good, demonstrate them in stereo or multichannel (to impress everybody). Choose the recordings carefully—they are a significant factor in how pleased you will be.

REFERENCES

Alexander, R. C. (1999). *The Inventor of Stereo—the Life and Works of Alan Dower Blumlein*, Focal Press, Oxford. ⏎

Augspurger, G. L. (1990). "Loudspeakers in Control Rooms and Living Rooms", *8th International Conference, Audio Eng. Soc*. ⏎

Bech, Søren (1989). "The Influence of Room Acoustics on Reproduced Sound, Part 1- Selection and Training of Subjects for Listening Tests", *87th Convention Audio Eng. Soc*., Paper 2850. ⏎

Benade, A. H. (1985). "From Instrument to Ear in a Room: Direct or via Recording", *J. Audio Eng. Soc*., vol. 33, pp. 218–233. ⏎

Benade, A. H. (1986). "Critique of Certain Concert-hall Design Criteria", Notes from an oral presentation at Anaheim, CA meeting of the Acoust. Soc. Am., December (personal communication). ⏎

Blumlein, A. (1933). British Patent No. 394 325 "Improvements in and Relating to Sound Transmission, Sound-recording and Sound-reproducing Systems" granted to Alan Blumlein and EMI, 1933. Reprinted in *J. Audio Eng. Soc*., 6, pp. 91–98, 130, 1958. ⏎

Briggs, G. A. (1958). *Loudspeakers*, 5th edition, Wharfedale Wireless Works, Ltd., Idle, Bradford, UK. ⏎

Choisel, S. and Wickelmaier, F. (2007). “Evaluation of Multichannel Reproduced Sound: Scaling Auditory Attributes Underlying Listener Preferences”, *J. Acoust. Soc. Am.*, vol. 121, pp. 388–400. ⏎

Clark, H., Dutton, G. and Vanderlyn, P. (1957). “The ‘Stereosonic’ Recording and Reproducing System”, Proc. IEE, Reprinted in: *J. Audio Eng. Soc.*, vol. 6, no. 2, pp. 102–117, 1958. ⏎

Davis, M. F. (1987). “Loudspeaker Systems with Optimized Wide-Listening-Area Imaging”, *J. Audio Eng. Soc.*, vol. 35, pp. 888–896. ⏎

Evans, W., Dyreby, J., Bech, S., Zielinski, S. and Rumsey, R. (2009). “Effects of Loudspeaker Directivity on Perceived Sound Quality—a Review of Existing Studies”, *126th Convention Audio Eng. Soc.*, Paper 7745. ⏎

Harvith, J. and Harvith, S. (1987). *Edison, Musicians and the Phonograph*, Greenwood Press, New York. ⏎

Hughes, R., Cox, T., Shirley, B. and Power, P. (2016). “The Room-in-room Effect and Its Influence on Perceived Room Size in Spatial Audio Reproduction”, *141st Convention Audio Eng. Soc.*, Preprint 9621. ⏎

ISO 389-1:2017Acoustics—Reference Zero for the Calibration of Audiometric Equipment. www.iso.org ⏎

King, R., Theriault, M. and Massenburg, G. (2023). “A Practical Approach to the Use of Center Channel in Immersive Music Production”, *155th Convention Audio Eng. Soc.*, Paper 168. ⏎

Meyer, J. (1978). *Acoustics and the Performance of Music*, Verlag das Musikinstrument, Frankfurt am Main. ⏎

Meyer, J. (1993). “The Sound of the Orchestra”, *J. Audio Eng. Soc.*, vol. 41, pp. 203–213. ⏎

Morton, D. (2000). *Off the Record*, Rutgers University Press, New Brunswick, NJ. ⏎

Moulton, D. (1995). “The Significance of Early High-Frequency Reflections from Loudspeakers in Listening Rooms”, *99th Convention, Audio Eng. Soc.*, Preprint 4094. ⏎

Moulton, D., Ferralli, M., Hebrock, S. and Pezzo, M. (1986). “The Localization of Phantom Images in an Omnidirectional Stereophonic

Loudspeaker System", *81st Convention, Audio Eng. Soc*., Preprint 2371 ⏎

Olive, S. E., Hess, S. M. and Devantier, A. (2008). "Comparison of Loudspeaker-Room Equalization Preference for Multichannel, Stereo, and Mono Reproductions: Are Listeners More Discriminating in Mono?", *Audio Eng. Soc. 124th Convention*, Preprint 7492. ⏎

Olson, H. F. (1957). "Acoustical Engineering", Van Nostrand, republished in 1991 by Professional Audio Journals, Inc. ⏎

Plenge, G. (1987). "On the Behavior of Listeners to Stereophonic Sound Reproduction and the Consequences for the Theory of Sound Perception in a Stereophonic Sound Field", *Audio Eng. Soc. 83rd Convention*, Preprint 2532. ⏎

Pulkki, V. (2001). "Coloration of Amplitude-panned Virtual Sources", *110th Convention, Audio Eng. Soc*., Preprint 5402. ⏎

Read, O. and Welsh, W. L. (1959). *From Tin Foil to Stereo*, Howard Sams, Indianapolis, IN. ⏎

Shirley, B. G., Kendrick, P. and Churchill, C. (2007). "The Effect of Stereo Crosstalk on Intelligibility: Comparison of a Phantom Stereo Image and a Central Loudspeaker Source", *J. Audio Eng. Soc*., vol. 55, pp. 852–863. ⏎

Snow, W. B. (1953). "Basic Principles of Stereophonic Sound", *J.S.M.P.T.E*., vol. 61, pp. 567–587. Reprinted in IRE Transactions-Audio, vol. AU-3, pp. 42–53, 1955. ⏎

Steinberg, J. C. and Snow, W. B. (1934). "Auditory Perspective—Physical Factors", *Electrical Engineering*, vol. 33, pp. 12–17. ⏎

Toole, F. E. (1970). "In-head Localization of Acoustic Images", *J. Acoust. Soc. Am*., vol. 48, pp. 943–949. ⏎

Toole, F. E. (1985). "Subjective Measurements of Loudspeaker Sound Quality and Listener Preferences", *J. Audio Eng. Soc*., vol. 33. pp. 2–31. ⏎

Toole, F. E. (1986). "Loudspeaker Measurements and Their Relationship to Listener Preferences", *J. Audio Eng. Soc*., vol. 34, pt. 1, pp. 227–235, pt. 2, pp. 323–348. ⏎

Vermeulen, R. (1956). “Stereo Reverberation”, *IRE Trans. Audio*, July–August, pp. 98–105. ⏎

Villchur, E. (1964a). “High Fidelity Measurements, Science or Chaos”, *Electronics World*, August. ⏎

Villchur, E. (1964b). “Techniques of Making Live-versus-Recorded Comparisons”, *Audio Magazine*, October. ⏎

Williams, M. (1987). “Unified Theory of Microphone Systems for Stereophonic Sound Recording”, *82nd Convention Audio Eng. Soc.*, Preprint 2466. ⏎

Williams, M. (2023). *Multichannel Microphone Array Design*. A book available only from the author at: mail@soundscot.com. ⏎

Wittek, H. and Theile, G. (2002). “The Recording Angle—Based on Localization Curves”, *112th Convention Audio Eng. Soc.*, Paper 5568. ⏎

Wittek, H. and Theile, G. (2017). “Development and Application of a Stereophonic Multichannel Recording Technique for 3D Audio and VR”, *143rd Convention Audio Eng. Soc.*, Paper 9869. ⏎

CHAPTER 5

Loudspeakers, Rooms, Recordings and Adaptation:Key Factors in What and How We Hear

Floyd Toole

DOI: 10.4324/9781003477495-6

All sound reproduction through loudspeakers involves a room as the final element in the sound delivery system. We know that listeners receive a direct sound from the loudspeaker and then many reflections from numerous directions at numerous later times. When examined, as was done in Section 2.4, we learn that the reflected sounds physically dominate what reaches the ears; collectively they are the loudest sounds. If the reader has not read that section, or time has passed, it might be useful to do a quick review of it.

The energy in the reflected sounds dominates measurements, but the first-arriving direct sound has a perceptual advantage. It determines sound localization—the precedence effect. It is a major factor in sound quality and

the reference against which later arriving reflected sounds are judged. Consequently, by simply adding the sounds together, as in measuring steady-state room curves, much important information is lost. Figure 1.5 illustrates the situation.

The reflected sounds are radiated from loudspeakers at large angles away from the on-axis direct sound, and because of that they are different in level and frequency response. What we hear, therefore cannot be anticipated by looking only at the loudspeaker or looking only at the room. The loudspeakers and room function as a system—together. This is apparent in the evolution of the spinorama measurement display processing described in Section 2.7.

Often, the room is regarded as a problem, and in some ways it is, but there are positive contributions as well, as were discussed in Chapter 4, moderating some problems specific to stereo. Finding the right balance is what we are after, and understanding the nature of the sound field inside a small room is important.

Finally, it is necessary to be alert to adaptation and its effects. Humans learn about, and become familiar with, various aspects of the environment, including how it sounds. In this case, it turns out that we have considerable ability to adapt to the reflective and resonance patterns in rooms so that we are able to make key decisions about the source of sound we are listening to.

5.1 ONE ROOM, TWO SOUND FIELDS—THE TRANSITION FREQUENCY

The curves displayed in Section 2.4 show that the right combination of anechoic measurements can closely anticipate events in typical rooms at middle and high frequencies, but at low frequencies the curves are lumpy and irregular. The acoustical explanation is the dominance of relatively isolated room resonances/modes and associated standing waves at low frequencies, and of a complex collection of overlapping modes and reflected sounds at high frequencies. As a result, as frequency increases it

becomes progressively less useful to think about regular patterns of standing waves in rooms, but rather to think in terms of irregular patterns of constructive and destructive acoustical interference caused by numerous reflections traveling in many directions.

Between the orderly low-frequency room resonances, and the different higher frequency reflection/ray-acoustical behavior, is a transition zone, the middle of which, in large rooms such as concert halls and auditoria, would be defined as the Schroeder frequency or, as Schroeder himself calls it, the cross-over frequency f_c (Schroeder, 1954, 1996):

$$f_c = 2000\sqrt{\frac{T}{V}}$$

where T is the reverberation time in seconds, and V is the volume of the enclosure in cubic meters. The multiplier constant changed from the original 4000 to 2000 in the 1996 paper, emphasizing that this is not a precision calculation.

Calculation of the Schroeder frequency assumes meaningful reverberation times, a strongly diffuse sound field and a volume unimpeded by large objects (furniture and people are large objects in small rooms). Small listening rooms and recording control rooms are too absorptive to have a meaningful diffuse sound field and a reverberation time—a *reflection decay time* would be a better description at middle and high frequencies and *resonance ringing time* at low frequencies.

Therefore, in small rooms, especially those with large furnishings, the strictly applied Schroeder frequency calculation is mismatched. The calculated value may be in error, as noted by Baskind and Polack (2000). For the room used in the measurements of Figure 2.6(c) the Schroeder frequency is 129 Hz (T = 0.32 s, V = 76.9 m^3). This appears to be on the low side, because the large undulations in the curves have not diminished. However, no matter how it is identified, or what it is called, the *transition region* is real, and it is necessary to take different approaches to dealing with acoustical phenomena above, within, and below it. It is unrealistic to think that room performance changes at a single frequency. Figure 5.1

illustrates the concepts, indicating that the sound sources, real or reproduced, instruments or loudspeakers, dominate what is heard over larger frequency ranges in large rooms.

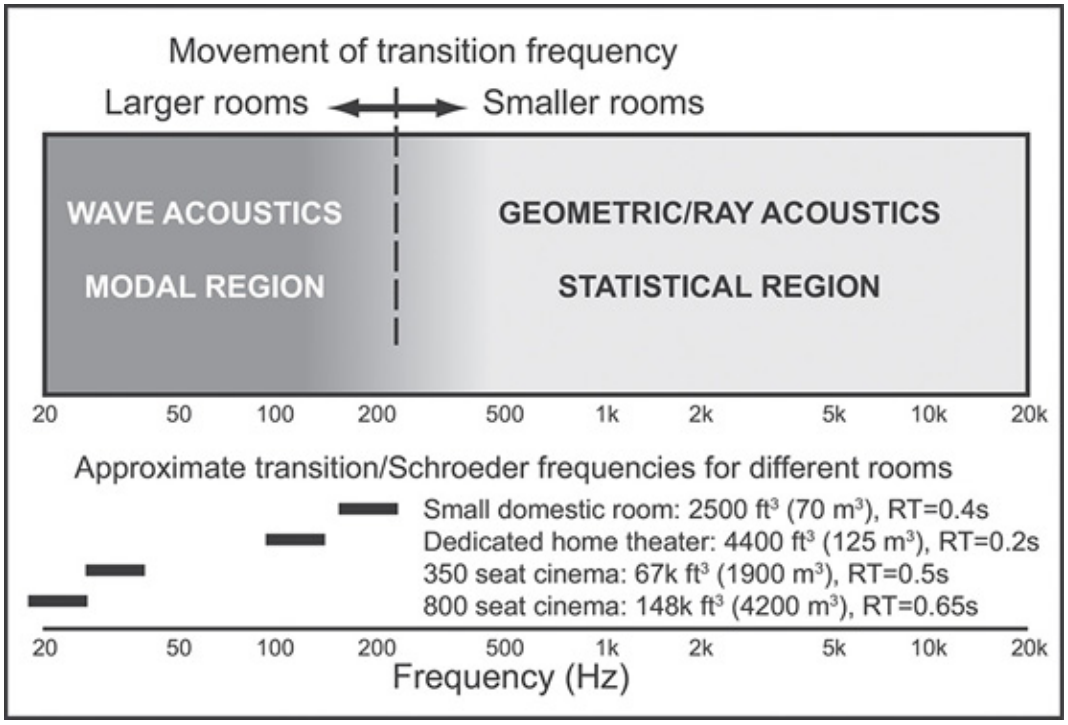


FIGURE 5.1 *An artistic interpretation of the transition between the low-frequency region dominated by room modes and the high-frequency region dominated by reflected sounds. Also shown is the effect of room size on the position of the transition region in the frequency domain. Although we talk of a frequency, the reality is a progressive change that takes place over a range of frequencies: a region.* ⏎

Figure 5.2 gives us more insight into this topic. Here, using an expanded frequency scale, high-resolution frequency-response measurements are shown from each of the five loudspeaker locations in a surround-sound system at the prime listening position. The room is dimensionally symmetrical, but differences in the curves reveal that it is not acoustically symmetrical. A door in one end wall causes it to flex more than the other one at certain frequencies, and a concrete wall behind, but not touching, one

of the side walls gives it more stiffness than the opposite wall. The result is asymmetry in the standing wave patterns. Five identical loudspeakers were arranged in the ITU-R BS.775–2 recommended arrangement and measurements were made at the listener's head location.

The standing waves (explained in Figure 9.18) cause huge variations at low frequencies, covering the full 40 dB range of the display. Above about 100 Hz the variations are attenuated and, above about 200 Hz, the closely spaced modes overlap, interfere with each other, and the curves settle down even more. Looking at the details, below about 200 Hz, in spite of some obvious variations related to the very different loudspeaker locations, one can see relatively independent resonant peaks at identifiable frequencies. The frequencies may not be an exact match to the calculated frequencies because the room boundaries were not infinitely stiff, as theory assumes.

Above 200 Hz, the pattern becomes much less orderly, and the peak-to-peak variations are smaller. Yet, an underlying trend is visible, including the step at 500 Hz, which is a characteristic of this loudspeaker, a subwoofer running without a low-pass filter. The real transition frequency for this room appears to be around 200 Hz. The calculated Schroeder (crossover) frequency (f_c) is 111 Hz ($T = 0.4$ s, $V = 128$ m^3), which seems to be too low. Substituting the original constant, 4000, into the Schroeder equation would yield a better answer for this small room.

The following chapters will deal separately with events above and below the transition frequency because they require different analysis and treatment. Out of respect for Schroeder's insight his name will continue to be used in this book, but transition frequency is also be used to indicate that we are dealing with small, relatively dead, listening rooms and not performance auditoriums. In neither case is it a precise calculation; it is an indicator of a transitional region.

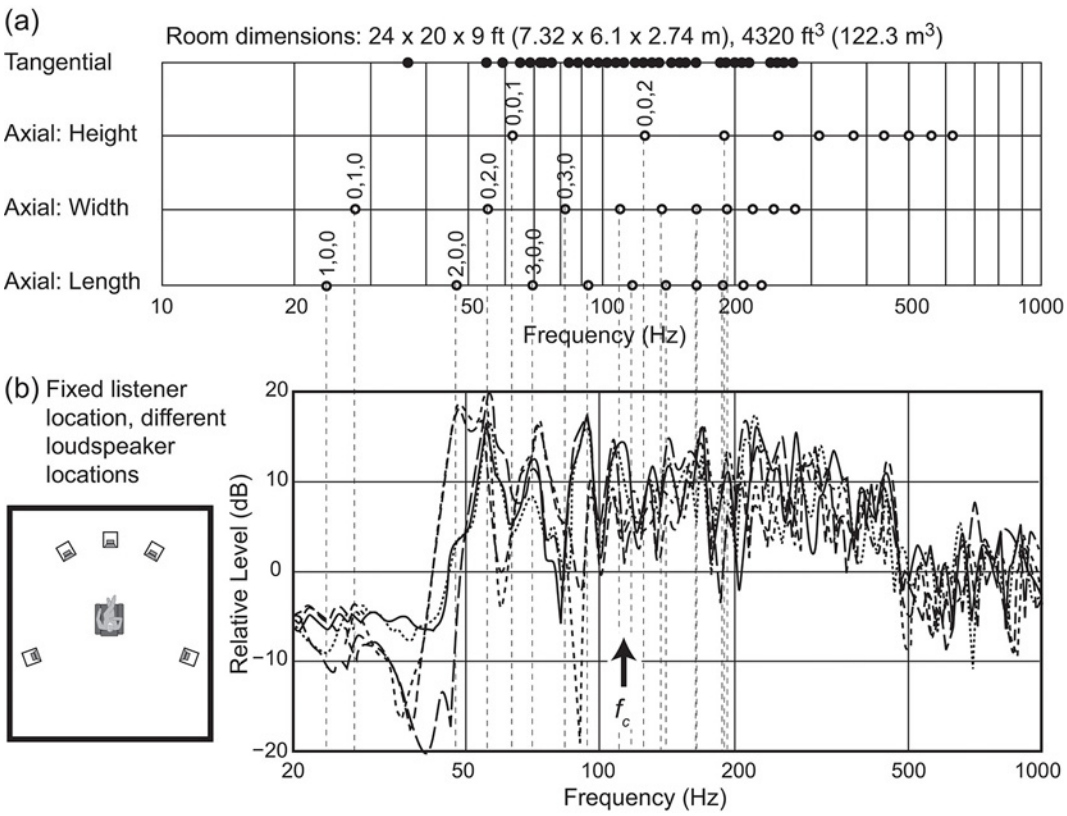


FIGURE 5.2 *(a) Axial and tangential modes calculated for a listening room set up with a listening arrangement according to ITU-R BS.775–2. (b) 1/20-octave steady-state measurements for each of the five loudspeakers, measured at the listening position. The Schroeder crossover frequency (f_c) is shown. Note the expanded frequency scale.*

5.2 TIMBRAL AND SPATIAL EFFECTS ATTRIBUTABLE TO ROOMS

Figure 5.3 is an attempt to summarize the perceptual effects that are experienced when loudspeakers energize a room. The direction and space effects are estimates based on comments in the literature. All will depend to some extent on the specific characteristics of loudspeakers, rooms and program. In the case of envelopment, cues are heard down to very low frequencies in large performance venues, but these are difficult to reproduce and hear in small listening rooms at frequencies below about 80 Hz.

5.3 RESONANCES NEED TO BE HEARD—OR NOT—IT DEPENDS

Section 1.5.1 introduced resonances as major elements determining the timbral identities of voices and musical instruments. These resonances need to be heard—this is the music, the instruments, and the voices. In contrast, resonances in loudspeakers reproducing voices and musical instruments modify those timbres, and they do it for all sounds reproduced—an obviously unwanted unchanging coloration.

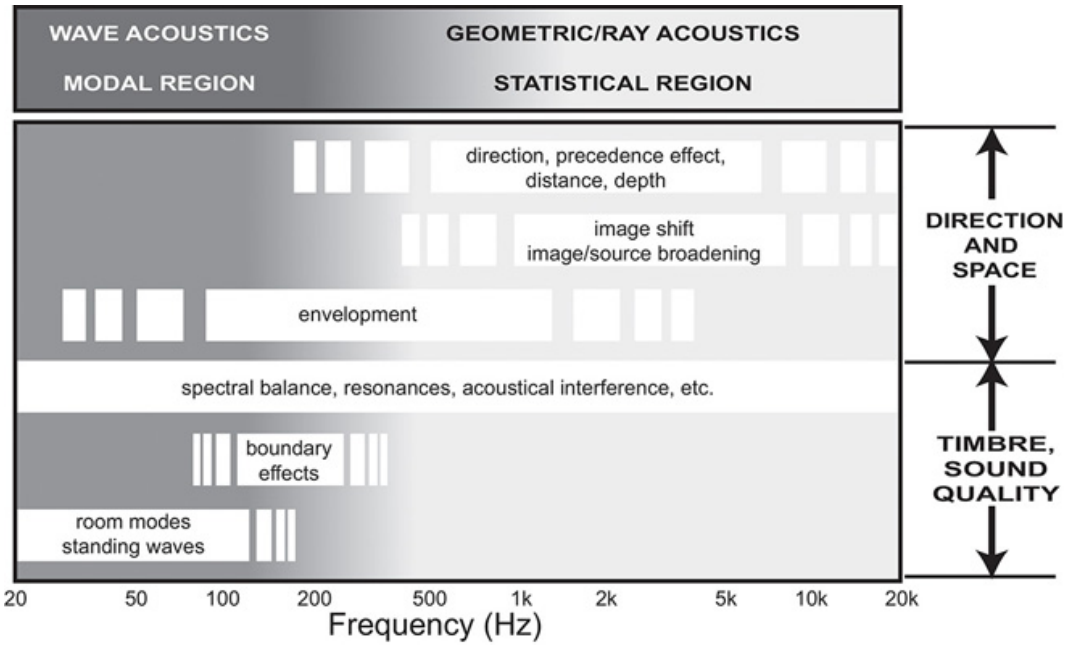


FIGURE 5.3 *The approximate frequency ranges over which various audible effects of reflected sound may be heard.* ⏎

In the earliest double-blind subjective evaluations of loudspeakers (Toole, 1985, 1986) it was apparent that audible resonances were responsible for most of the lower sound quality ratings. Listeners could hear the added colorations and preferred more neutral loudspeakers. These are resonances we don't want in loudspeakers, but they need to be revealed in subjective evaluations. In Figure 5.3 the word "resonances" appears in the middle of a bar that embraces all frequencies.

Figure 5.2 shows that in small rooms, low-frequency resonances are significant problems. They are different in different rooms, and different for every loudspeaker/listener arrangement within a room. The loudspeakers provide low-frequency energy, but the room is the dominant factor in determining the quality of the sound below the transition frequency. It is a problem that must be addressed by every user. Chapter 14 will deal with it in detail.

Above the transition frequency it is resonances that give the voices and instruments we listen to their distinctive characters, and resonances in the loudspeakers that degrade those characters. Obviously, loudspeakers should be neutral, aurally transparent reproducers. It therefore is important to know:

1. How to recognize the presence of resonances in subjective loudspeaker evaluations.
2. How to recognize the presence of resonances in loudspeaker measurements.
3. How to evaluate how audible those resonances are.
4. The extent to which room reflections affect the audibility of resonances, and therefore the sound quality of reproduced sound.

Item 1: The earliest listening tests employed an experimental method that turned out to be very effective at exposing audible differences among a small group of loudspeakers (Section 1.3.1). Those tests were conducted in mono. When similar evaluations were done in stereo, it was found that listeners were more forgiving of flaws in lower-rated loudspeakers; see Section 4.4 for details. When expanded to include the 5-channel multichannel format, listeners were even more forgiving of flaws in lower-rated loudspeakers; see Section 4.5. The top-rated loudspeakers retained that position in all listening formats. It is reasonable to assume that those flaws were resonances, and that the sound field complications of additional channels modified the audibility of those resonances. But were the relevant

complications simply due to the additional loudspeakers in the room, or to the additional sound field information in the recordings being listened to? Obviously, both are present simultaneously.

Item 2: The early measurement method, involving comprehensive anechoic frequency responses, revealed that resonances were major factors in sound quality ratings, and that they could be identified as repeated patterns in the families of frequency-response curves and discontinuities in the phase-response curves (Section 2.3).

Item 3: The following section explains research that identified the audible thresholds of resonances as visible deviations in frequency response curves.

Item 4: This is unfinished work, but much is known. Some of it is suggested in the item 1 commentary, and some of it has to do with familiarizing ourselves with our listening environment, which we can call adaptation. Adaptation is at work in daily auditory perceptions, allowing us to "normalize" the acoustical environment within which we listen. Imagine a scenario in which you and a colleague are conversing while walking down the street (a "dead" space). You enter a building foyer (a large reverberant space), take the elevator (a tiny space) and walk down a corridor (a distinctive reverberant space) into an open plan office (a relatively dead, semi-reflective space). While you are conversing, your colleague's voice is a constant factor in spite of the enormous changes in the acoustical environment. It is the same voice in different acoustical spaces. The only relatively constant factor in all of these situations is the direct sound, combined with the cognitive correlation of visual cues with what it being heard. It all happens subconsciously, and very quickly. This example is of a single voice, a mono signal, but there are parallel experiences in live musical performances in different venues. The acoustics of the venues do not appear to corrupt the essential characters of the voices and instruments, yet listeners are aware of the acoustics of the venues. Keep these thoughts in mind as you read on. Adaptation is always with us.

5.3.1 Measuring and Hearing Resonances

Resonances concentrate energy at specific frequencies and exhibit their presence in both the frequency and time domain. Tap the rim of a wine glass while holding it by the stem and it rings with a distinctive pitch with a long overhang—ringing. This is a high-Q resonance. Do it again with the holding hand just touching the bottom of the bowl. It rings, but with less overhang. The hand is damping the resonating glass bowl. This is a medium-Q resonance. Do it a third time holding the glass by the bowl itself. There may be a hint of a pitch, but the event is very brief. This is a low-Q, well-damped resonance. Instinctively, one would think that high-Q resonances would be the major audible problem in loudspeakers, but instinct is wrong.

Simple resonances are found in electronic, acoustical and mechanical devices. It is often useful to think of them as comprising "lumped elements" in a resonating system: a mass component (inductance), a spring/compliance (capacitance) component and a damping (resistance) component. The mass and compliance determine the frequency of the resonance. The damping determines the amount of energy loss in the system, which defines the quality factor, or Q. This determines the bandwidth of the resonance in the frequency domain and the duration of the buildup and decay of energy in the time domain. Not all resonances are of the lumped element variety. They can occur in the radial and circumferential wave motion in loudspeaker diaphragms (and wine glasses). When reflections from the edge surrounds reinforce the incoming waves from the voice coil, there is an aberration in the frequency response. And there are other mechanisms, including panel resonances in enclosures, all of which add to the sound radiated by loudspeaker systems.

High-Q resonances have a small footprint in the frequency domain (a narrow, sharpish spike in a frequency response) and a long buildup and decay in the time domain. As Q falls, the resonance gets wider in the frequency domain, and time-domain events happen more quickly. Eventually, the frequency response gets to be "flat", there is no evidence of a resonance and the time-domain misbehavior disappears. So, it is inevitable that all systems designed to minimize resonances have flattish,

smooth frequency responses. Virtually all electronics meet this objective, and the best loudspeakers are close.

Resonances in loudspeakers are unwanted, and it is one of the prime challenges of loudspeaker designers to eliminate them because they radiate throughout the room. Toole and Olive (1988) describes the background history of research in this area and adds considerably to it. The goal is to define in technical, measurable terms reliable evidence of audible resonances in loudspeakers. What was found confirmed existing evidence that it is the low-Q resonances that are most frequently stimulated by music, making them most easily heard by our auditory systems.

The nature of the program material is a significant factor. Büchlein (1962), Toole and Olive (1988) and Olive (1994) found simple sounds, solo vocal and instrumental recordings to be the least revealing of differences between loudspeakers. Although film and broadcast people place great importance on voices, for good reason, they happen to be among the sounds that are least revealing of imperfections in sound reproducing systems. High speech intelligibility is possible through systems that are significantly flawed when judged from a timbre/sound quality perspective. Many audiophile demonstrations use uncomplicated recordings of solo voices or instruments with simple accompaniments. They sound good, eliciting flattering comments, possibly because they are not revealing resonances in the loudspeakers being demonstrated. The most useful recordings have broadband, dense spectra as appear in recordings with high production values, some reverberation and complex orchestrations. Pink noise is the most revealing signal. All we need to know is that it should not exhibit recognizable pitch(es).

Reflections also increase sensitivity to medium- and low-Q resonances, apparently as repetitions of the direct sound, offering the hearing system additional “looks” to gain more information. Toole and Olive (1988) reported that the repetitions (reflections) necessary to increase our sensitivity to resonances can be either in the recorded sounds themselves or contributed by the listening room. When listening through headphones or in a dominant direct sound field (a very dead room) we are less sensitive to

low- and middle-Q resonances. This would appear to be a factor in subjective assessments of headphone sound quality, which will be discussed in Chapter 15. It is also a factor in recording control rooms, many of which strive to eliminate reflections, and where close (near-field) listening is common. These are situations in which the recording engineers will hear less timbral enrichment than typical consumers. Adding some electronic reverberation "sweetens/enriches" the sound, though, because the desired repetitions can be in the listening room or in the source material.

Figure 5.4 shows some resonances precisely at the thresholds of detectability for pink noise, the most revealing signal. The key idea being communicated is that neither the amplitude of the peak in the frequency response nor the duration of ringing is a reliable clue to the audibility of the resonances, at least above the transition frequency. It is counterintuitive. Adding damping reduces the Q, and therefore the ringing, and the amplitude of the peak in the frequency response, all of which would seem to be good. But the peak is now much wider and is therefore more frequently energized by spectral components of the program. This means that it will be more often heard; the threshold ends up being lower. The low-Q (Q = 1), well-damped resonance is easily heard, as are all wide-bandwidth equalization changes.

In contrast, the Q = 50 resonance is energized only when a musical component is precisely at the right frequency and remains there long enough to allow it to accumulate energy—to build up—which is easily seen in the center column of data. The amplitude of the ringing is of the same order as the tone burst itself, but it happens relatively infrequently. When the right frequency appears in music, as happens occasionally with voice formants or instrument spectral components at the same frequency as a resonance, it is very audible and very annoying. So, in long-term listening, the probability of hearing a high-Q resonance is low, but if circumstances are right, the problem can be severe. Long-term listening may be required before such a problem is revealed, but a measurement shows it immediately. Clearly *all* resonances should be minimized.

The key question: “Do we hear the peak in the frequency response or the ringing or both?”

We may hear some of both, but for frequencies above about 200 Hz, the *measurable evidence* of audibility appears to be the spectral peak. The audibility of resonances is best described by the anechoic frequency response curve. This topic is discussed in more detail in Toole and Olive, 1988. At low bass frequencies the time-domain misbehavior exists over longer durations, and there is the potential of hearing both the spectral feature and the ringing. This will be discussed in Chapter 14.

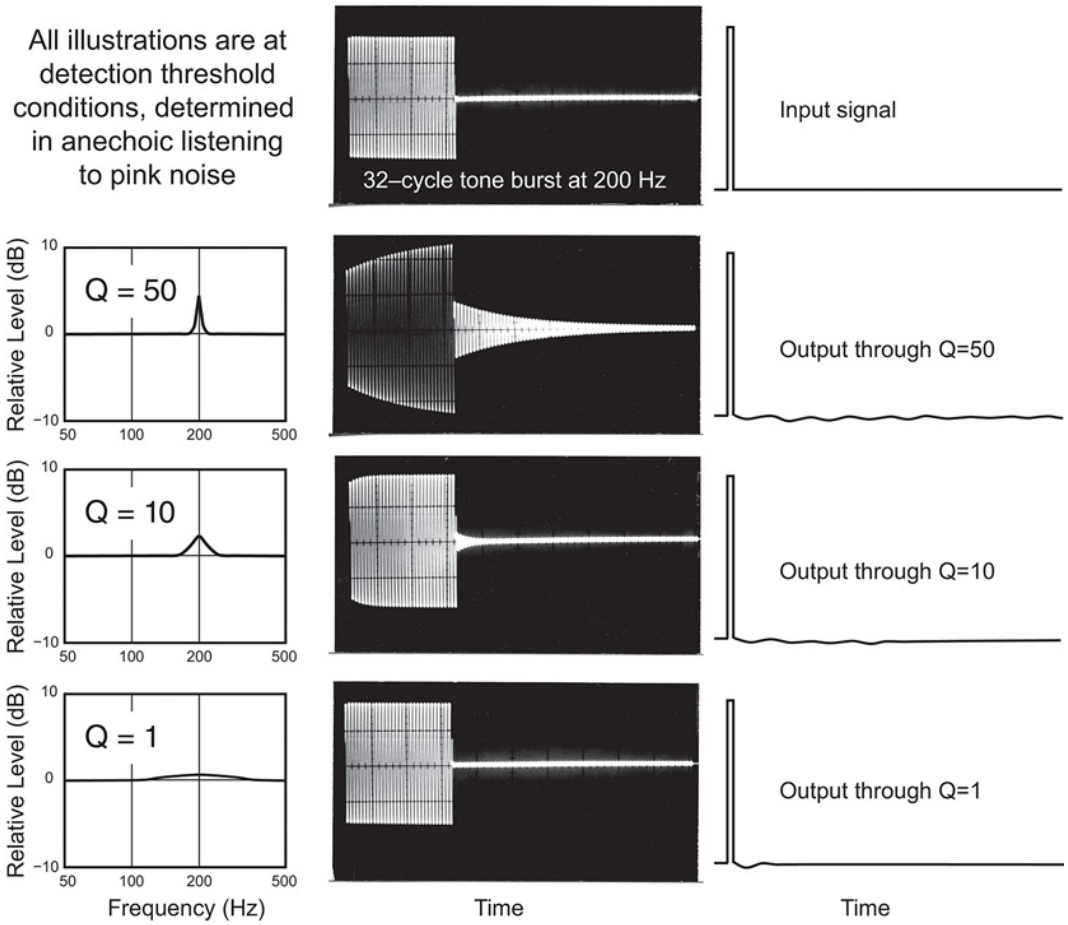


FIGURE 5.4 *On the left are shown frequency responses of resonances when they are at the threshold of detection when listening to pink noise, the most revealing of all program material. The high-Q (Q = 50) resonance is just audible when the amplitude of the narrow peak is about 5 dB. The medium-Q (Q =*

10) resonance is just audible when the amplitude of the wider peak is around 2 to 3 dB. The low-Q (Q = 1) resonance is just audible when the amplitude of the broad hump is about 0.5 dB. In the center column are oscilloscope pictures of how these resonances respond to a tone burst at the center frequency of the resonance. The high-Q resonance takes time to accumulate energy and a similar amount of time to release it in the form of ringing. The medium-Q resonance rises to maximum amplitude quickly and decays quickly. The low-Q resonance rises and decays almost instantaneously. The right column shows how these resonances respond to brief impulses. Again, the high-Q resonance exhibits the most sustained ringing. Figure 4.10 in the third edition of this book, adapted from Toole and Olive (1988). ⏎

All resonances in the playback devices need to be attenuated as much as possible. Neutrality is the goal. Flat and smooth direct-sound anechoic frequency responses are desirable.

The right-hand column shows a very different situation for impulses. Because the spectrum of a brief transient is very broadband, with little energy at any one frequency, there is ringing, but at a very much lower amplitude than the peak of the impulse. Forward temporal masking (a perceptual phenomenon) will reduce the perception of the ringing even farther.

Figure 5.5 shows examples of deviations from flat for high-Q (Q = 50), medium-Q (Q = 10) and low-Q (Q = 1) resonances at three frequencies when they were adjusted to the audible threshold levels using pink noise in an anechoic chamber, and for the 200 Hz resonances detected when listening to typical close-miked, low-reverberation pop and jazz. In these simple recordings the absence of repetition (reflections/reverberation in the

mix) and possibly a lack of spectral complexity has reduced the listener's ability to hear the resonances.

The top frequency response displays just audible high-Q (Q = 50) resonances for pink noise, auditioned in an anechoic chamber. At 200 Hz the peak was about 4 dB, at 1 kHz and 5 kHz the peak was about 1.5 dB. With close-miked, low-reverberation pop/jazz music the 200 Hz resonance peak was about 10 dB. The middle curve shows just detectable medium-Q (Q = 10) resonances. For pink noise the 200 Hz peak was about 2 dB, the 1 kHz peak was about 1 dB and the 5 kHz peak was about 0.5 dB. With pop/jazz the 200 Hz peak went up to about 6 dB. The bottom curve shows detection thresholds for low-Q (Q = 1) resonances. For pink noise the 200 Hz peak was about 1 dB, and the 1 and 5 kHz peaks were around 0.25 dB. With pop/jazz the 200 Hz resonance peak went up to about 2.5 dB.

The detection thresholds for low-Q resonances are astonishingly low, explaining why small changes in spectral balance or tone-control adjustments are easily heard. These probably represent the most demanding criteria for identifying audible resonances in frequency response curves.

Also shown is the frequently quoted ±3 dB tolerance for frequency response curves. This can be seen to be too tolerant of medium- and low-Q resonances and unnecessarily restrictive of some high-Q resonances. Performing within the limits of this tolerance is an assurance of nothing in terms of sound quality, but marketing departments have been faithful users of it in specification sheets.

The evidence of audible resonances shown in Figures 5.4 and 5.5 should be performance criteria for the highest-quality loudspeakers. But they are useful data in designing loudspeakers at any price level, because they enable "value engineering", i.e. providing maximum sound quality at the lowest price. Figure 2.7 illustrates the result: good-sounding loudspeakers at low prices, crowded together in the upper left-hand part of the distribution. Price alone is no assurance of high sound quality.

Summarizing

1. The amplitudes of the resonances shown in the frequency responses are the anechoic *steady-state* measured changes in the playback system caused by the presence of the resonances that have been adjusted to the detection-threshold level while listening to different kinds of program in an anechoic environment. The fact that the resonant peaks are higher for the chosen pop/jazz example reflects the fact that the program material exhibits a lower probability of exciting the resonance than pink noise, a spectrally dense, steady-state signal. Not shown, but interpretable from the Fryer (1975, 1977) data, shown in Toole and Olive (1988), is the fact that symphonic music, which is both spectrally complex and reverberant, exhibits thresholds that are between the pink noise and pop/jazz thresholds shown in Figure 5.5 at 200 Hz.
2. High-resolution measurements are necessary to reveal high-Q features in the frequency responses. Because they exist at all frequencies, including the lower-frequency regions, it means that anechoic or digitally simulated anechoic measurements are necessary to reveal them.

   Any tolerance applied to a frequency-response curve needs to take into consideration the bandwidth/Q of the deviations that are being described. The conventional ±3 dB tolerance can be misleading unless one is able to see the curve(s) that are being described by this generous tolerance number. Such specifications are useful only if the tolerance is small and the measurement resolution is high, e.g. 1/20-octave. At ±1.5 dB sound quality is promising, but seeing the curves is more revealing. Some widely used international standards also use simple tolerances in 1/3-octave band measurements, the low resolution of which further hides resonant misbehavior (e.g. ITU-R BS.1116-3, 2015).
3. Finally, all these threshold determinations were done in anechoic listening conditions. As mentioned, these thresholds may be even

lower when listening in reflective rooms. However, if the thresholds were determined using signals incorporating significant repetitions (e.g. reverberation), the effect of the listening environment is minimal.

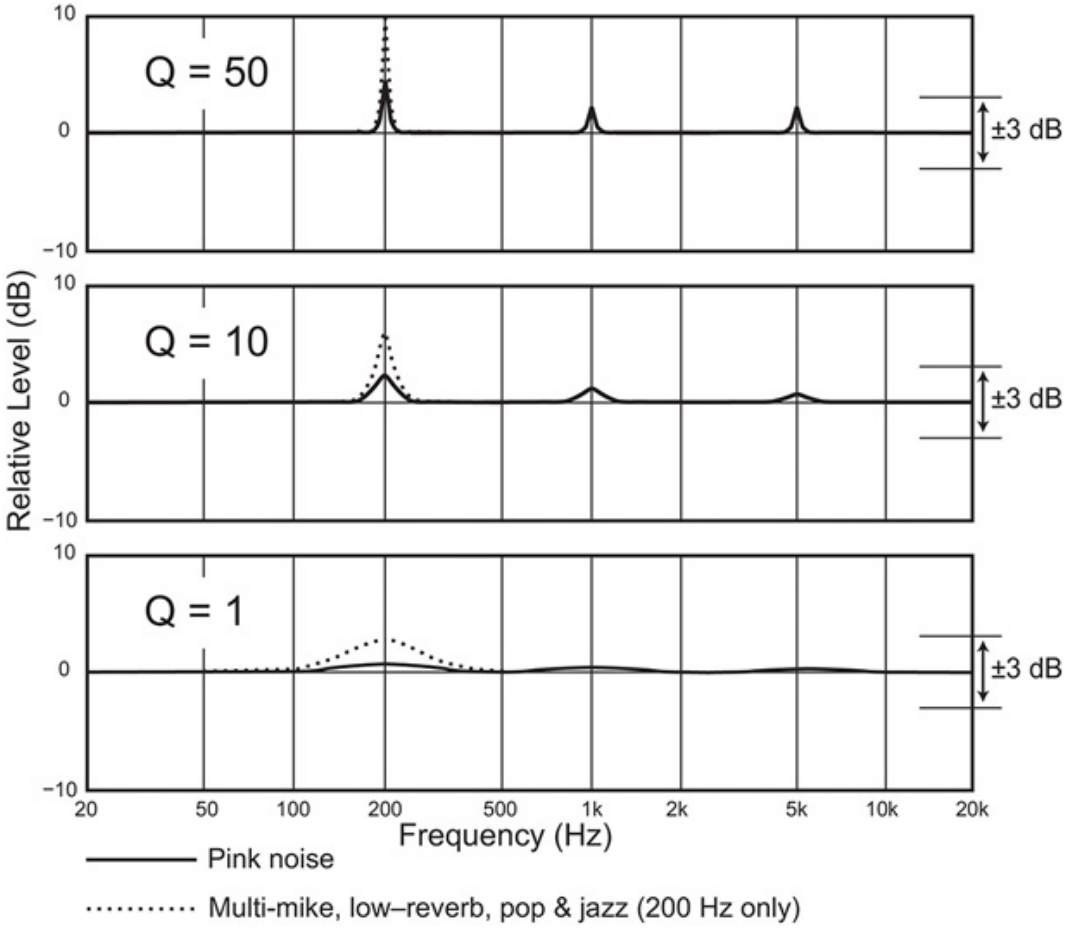


FIGURE 5.5 *Deviations from flat caused by resonances of three different Qs that were just detectable when listening to pink noise (all three frequencies) and to close-miked, low-reverberation pop/jazz recordings (200 Hz only). Adapted from Toole and Olive (1988).* ⏎

5.3.2 Finding and Fixing Resonances: Different Methods—Different Results

The fact that loudspeaker transducers are *minimum-phase* devices means that parametric equalization can be used to address resonances in them.

Reducing the amplitude of a resonance peak reduces the audibility of the resonance and damps the time-domain ringing. In technical terms minimum-phase systems are “invertible”. This is a powerful argument for loudspeakers with dedicated electronics, including DSP. All of this evidence confirms the logic of designing loudspeakers to have smooth and flat anechoic on-axis frequency responses, which is what listeners in double-blind multiple-comparison listening tests have told us for decades (Toole, 1985, 1986, Section 2.3.1).

The challenge is to identify resonances in the fluctuating frequency response curves that also contain peaks and dips associated with acoustical interference, a non-minimum-phase problem that is not correctable by parametric equalization. Fortunately, interference is a much less audible problem because in most instances it is different for every off-axis angle and is averaged out in our perceptions in reflective rooms. Resonances tend to radiate uniformly over large angular ranges. They are visible as repeated features in families of anechoic on- and off-axis frequency responses, and audible in the direct and reflected sounds by everyone in the room.

In order to properly equalize a resonance in a loudspeaker, the center frequency, Q and amplitude must be known. This requires a moderate amount of frequency resolution in the measurements and a parametric equalization filter that can be appropriately tuned in center frequency and Q.

Figure 5.6 shows anechoic measurements on a loudspeaker before and after it has been equalized in different ways as shown in Olive et al. (2008).

Resonances radiate on axis and can be seen with some consistency in spatially averaged far off-axis measurements as well. Acoustical interference effects change with microphone location. Therefore, they are averaged out of curves that are combinations of several measurements at different angles, as in the early reflections and sound power curves. The bumps that are consistent through the family of curves are resonances. In Figure 5.6(a) it can be seen that there are four strong resonances. Keen eyes may find others, but they are so low in amplitude that they are not audible problems. The most significant resonance is the broad low-Q hump around

3 kHz. It can be seen that this bump is several times higher than the detection threshold for a Q = 1 resonance, the closest match. The other medium-Q resonances in (a) are close to the anechoic detection thresholds, but in a reflective room the thresholds will be even lower, so these may be audible in the right kind of program.

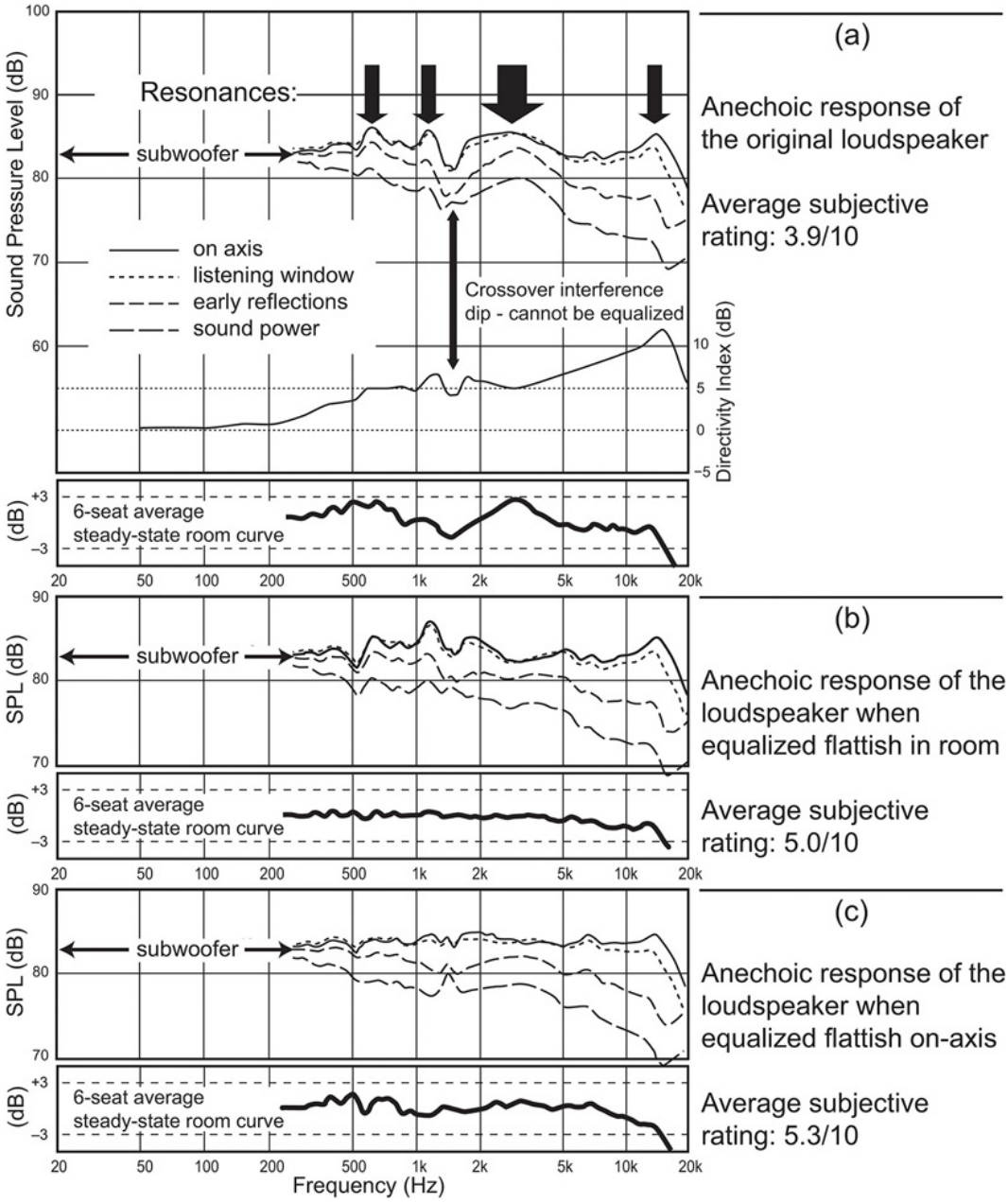


FIGURE 5.6 *Spinorama measurements on a loudspeaker in an anechoic chamber and as a steady-state six-seat average in a*

listening room. Below 200 Hz, all situations used the same subwoofer system to normalize the bass performance, which is very important because variations in bass are major influences on sound quality evaluations (Olive, 2004, 2004b). The loudspeaker has been equalized above about 400 Hz in two different ways: a smooth, slightly declining room curve and a smooth, flattish anechoic on-axis curve. Here it is possible to see the performance of the loudspeaker measured in both ways for each circumstance. These data have been extracted from Olive et al. (2008), and detailed information about the tests may be found there. The subjective ratings shown at the right are averages of all three listening modes. ⏎

The loudspeaker in its original state is an undistinguished product, not rating very high in the normal monophonic double-blind evaluations done at Harman. Here it has been used as a medium for evaluating different ways to try to improve its sound using equalization. The two methods shown here are the obvious ones:

1. Equalize a spatially averaged room curve to be reasonably flat and smooth.
2. Equalize the loudspeaker to have a reasonably flat and smooth on-axis frequency response in an anechoic chamber.

The loudspeaker seemed like a good candidate for equalization because of its generally well-behaved directivity. It will be noted, however, that there is an interference dip around 1.5 kHz caused by a flaw in the crossover. Because it is acoustical interference, it is averaged out with increasing spatial averaging and almost disappears in the sound power curve. Because of this it shows up in the directivity index. This is usually a non-minimum-phase phenomenon, and in those cases should not be equalized—but, as will

be seen, it was in these experiments that employed an automatic EQ algorithm.

The special, and very revealing, feature of these data is that we get to see the performance of the loudspeaker in comprehensive anechoic data, together with its performance in a room as revealed by the steady-state frequency response, averaged over several listening locations. We also get to relate both of these data sets to double-blind subjective evaluations. This is a powerful experiment, of which more need to be done. The problem is that few places have the facilities, budget and motivation to do it.

Based on these results it appears that any form of equalization was beneficial, elevating the subjective ratings from an unimpressive 3.9/10 for the original loudspeaker to 5.0/10 for the in-room equalized version and 5.3/10 for the anechoic on-axis equalized version. It is clear that the medium-Q resonances do not show up in the room curve in (a); therefore, when the room performance was equalized as in (b), only the low-Q resonance was attenuated. The subjective ratings went up substantially because that was the most serious problem. The medium-Q resonances were still there, only slightly changed.

What we have learned:

- In the anechoic data all of the problems are on display, and all are addressed by the on-axis equalization. The result is the highest subjective rating.
- The anechoic chamber equalization improved the loudspeaker, suggesting that if one had started with a well-designed loudspeaker, no equalization would be necessary above the transition frequency. The motivation is clearly for manufacturers to build timbrally neutral loudspeakers, and for customers to seek out and purchase such loudspeakers.
- Consumers with a technical inclination and access to accurate *anechoic* data on the loudspeaker can apply parametric equalization to

improve an imperfect loudspeaker, as done in this experiment and as is done in all active loudspeakers.

- Room curves do not reveal all loudspeaker problems, and in-room equalization to achieve an immaculately smooth room curve *above the transition frequency* is not sufficient.
- In all cases the room resonance problems *below the transition frequency* remain, and for this only, the steady-state room curve is a good measure of the problem and equalization and/or multiple subwoofer solutions can be very helpful, all of which will be discussed in Chapters 13 and 14.

Not to be overlooked is the fact that in (c) the automatic equalization process also filled the interference dip in the on-axis curve. This resulted in the addition of a resonance to the loudspeaker, which now shows up in the sound power curve. We will never know if the subjective scores would have been different if that had not happened. Human interpretation of the spinorama measurements is advisable; automatic equalization is not intelligent equalization.

A major purpose of these tests was to compare subjective rating performances when listening in mono, stereo and 5-channel multichannel modes. Figure 5.7 shows the result, and it is very clear that adding more active channels to a listening experience reduces one's ability to discern details of loudspeaker performance—but the least flawed loudspeaker (the anechoically equalized version) was recognized in all playback modes, albeit with statistical significance only in mono mode. Figure 5.7(b) shows evidence of subjective scale elasticity, illustrated in Figure 1.4. In mono mode, where the contrast in sound quality was strongly heard, the two equalized versions were similarly rated at the top of the scale. In stereo and multichannel the sounds were judged to be more similar, and range of ratings is progressively reduced, but the preference for the top-rated loudspeaker remains as a visible trend.

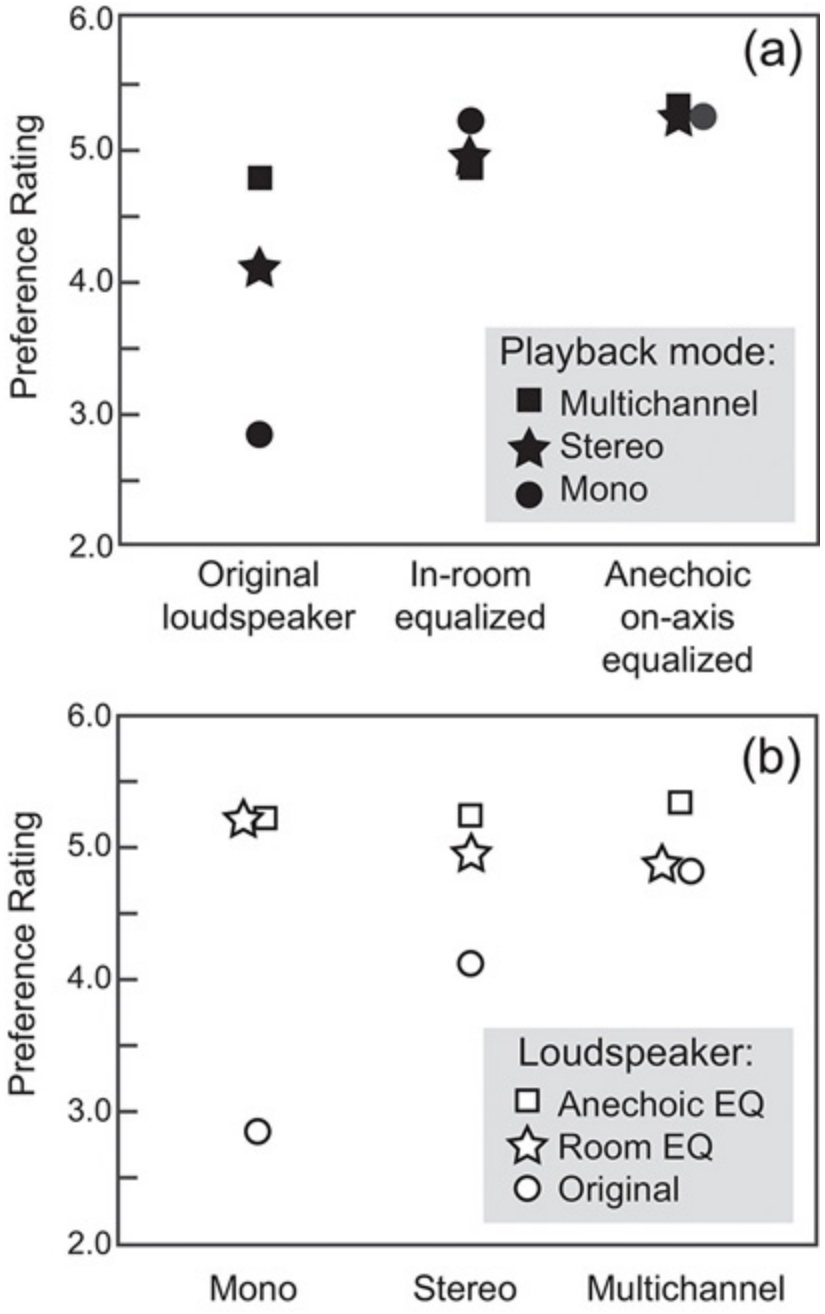


FIGURE 5.7 *The subjective ratings of the loudspeaker auditioned in three playback modes: mono, stereo and multichannel presented in two ways. (a) is from Olive et al. (2008). (b) was created by the author from that data to show the progression of rating changes as playback mode is altered from mono to stereo to multichannel. The only statistically significant result is in the mono test, but clear trends are visible. The original*

loudspeaker can be seen to improve its rating through the addition of channels. The rating of the in-room equalized version fell slightly as the range of perceived sound quality reduced with the addition of channels and listeners focused more on small differences. The anechoically equalized version maintained the highest rating throughout, even rising slightly. ⏎

What is happening? A quick scan down the families of curves in Figure 5.6 reveals that the technically best loudspeaker (c), anechoically equalized to have a very smooth, flattish on-axis frequency response, got very high ratings in all playback formats. The second-best, in-room equalized loudspeaker (b) also was awarded high scores in all modes, trending lower as the range of perceived sound quality was reduced. The original loudspeaker (a) had clearly visible and audible resonances when auditioned in mono, and was appropriately given low ratings, but in stereo the resonances were apparently more difficult to hear, and it was given higher ratings. In multichannel, it was judged to be not statistically different from the two equalized versions—the problem resonances were apparently not noticed.

With increasing numbers of channels the resonances which are highly visible, and audible in mono, are progressively less well heard. With the exception of the major problem in this loudspeaker, a low-Q resonance, they are also not accurately revealed in the steady-state room curves. There is more to this story—as will be seen, it is not just the number of channels, it is also the program content of the channels.

5.4 ROOM EFFECTS, LISTENING METHODS AND ADAPTATION: AN IMPORTANT EXPERIMENT

One of the last experiments done at the NRCC before the research was terminated occurred after I left to join Harman International in California. It utilized the suite of adjustable rooms designed for evaluating the effects of different room sizes and shapes on perceptions of sound quality. Earlier

tests had confirmed the fact that the location of a loudspeaker in a room affects its sound quality rating, and that low-frequency room resonances were a large part of the effect (Olive et al., 1992, 1994). This was expected, but results from later work was very instructive. Schuck et al. (1993) published detailed results on a collection of experiments. I was listed as an author, but my role was as an enabler of the experiments, not an active participant, as I was at Harman International in California at the time. One of the experiments was the focus of Olive et al. (1995). Both accounts are worth reading. These experiments were challenging to perform, and the material and manpower cost was considerable. They deserve our attention.

The publications give very detailed results of an elaborate test in which three loudspeakers were subjectively evaluated in four different rooms in double-blind "live" listening tests, and in binaurally recorded renderings of the same and another different test through headphones. There is massive statistical analysis of the many factor interactions in the experiments. It is not always clear how they affect the results, but in the end it is comforting to see the high statistical significance of a few differences that matter greatly.

Figure 5.8 shows the layouts of the four rooms used in the evaluations. Loudspeakers were hidden from view by acoustically transparent screens in rooms that were furnished in domestic fashion and were acoustically similar to typical domestic listening rooms. The papers contain detailed descriptions. Each of the three loudspeakers was auditioned in each of the three positions in each room. This helps to average out the low-frequency room resonance and other position-related interactions with the rooms.

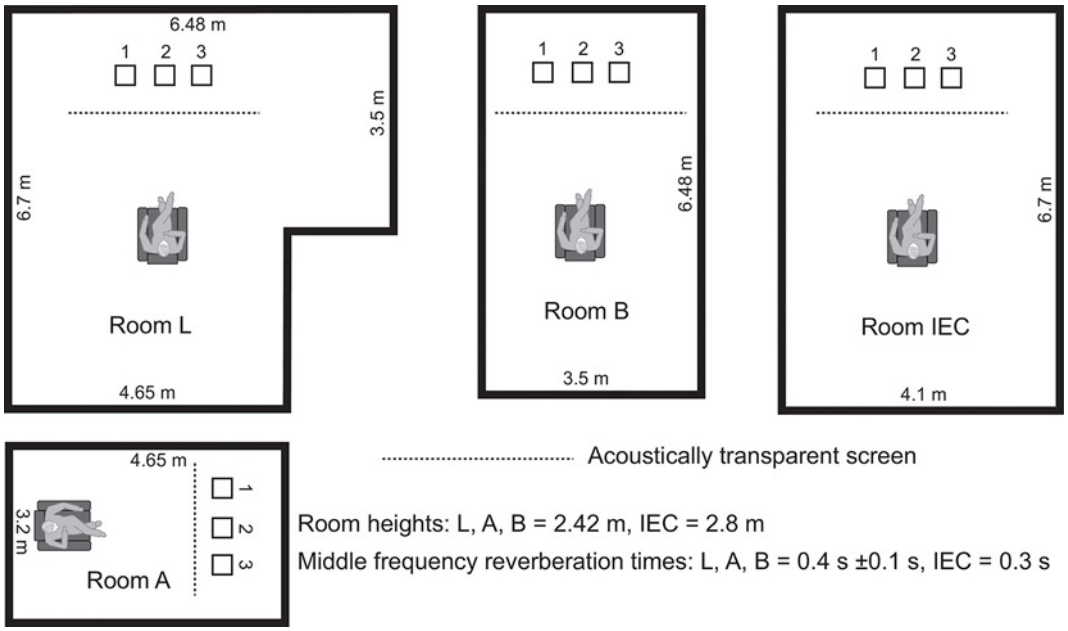


FIGURE 5.8 *Four listening rooms used in the experiments. There is one "L" shaped room and three rectangular rooms of very different sizes and proportions. In each room, listeners auditioned three loudspeakers in a very similar arrangement with respect to the listener, hidden behind an acoustically transparent screen. The loudspeakers were evaluated in each of the three locations, using three different programs, by 20 listeners. Binaural recordings were made for subsequent headphone reproduction. (Based on Schuck et al., 1993, figure 19.)* ⏎

Experiment One: In the "live" test, listeners completed the evaluation of the three loudspeakers in one room before moving to the next one. The loudspeakers were all forward-firing cone/dome configurations, having similar directivities and good overall performance.

Binaural recordings were made of each loudspeaker in each location in each room, and the tests were repeated, but this time with listeners hearing all of the sounds through calibrated headphones. All tests were double blind. In each room, three loudspeakers were evaluated in three locations for each of three programs. The whole process was repeated, resulting in 54

ratings for each of the 20 listeners. The result from a statistical perspective was that:

- "Loudspeaker" as a factor in the results, was highly significant ($p \leq 0.05$).
- "Room" was not a significant factor.
- Preference ratings determined in the live and binaural tests were essentially the same.

The dominating variable in this experiment were the three loudspeakers being instantly compared with each other. A possible interpretation is that the listeners became familiar with—adapted to—the room they were in and, as a result, were able to judge the relative merits of the loudspeakers. Obviously, this adaptation was also influenced by the different loudspeaker locations.

It is important that the binaural version of the live test yielded essentially the same loudspeaker preferences because this enabled the very revealing test that followed.

Experiment Two: Using the same binaural recordings that closely replicated the results of the live listening tests, another experiment was conducted. In this, the loudspeakers were compared with themselves and each other when located in each of the loudspeaker positions in each of the four rooms. Thus, in this experiment, the sound of the room was combined with the sound of the loudspeaker in randomized presentations that did not permit listeners to become familiar with the listening space. The result was that:

- "Room" became the highly significant variable ($p \leq 0.001$).
- "Loudspeaker" was not a significant factor.

Listeners were not able to normalize the rooms, and the differences among the rooms were perceptually more significant than the differences among the loudspeakers. The rooms dominated the results.

Pike et al. (2013) confirmed these basic findings using different loudspeakers, different rooms, and male speech as a signal. There is incomplete understanding of the perceptual mechanisms underlying the phenomenon, but there is no doubt that timbral adaptation occurs, and it applies to both loudspeakers and listening rooms. This appears to be related to the spectral compensation effect noted by Watkins (1991, 1999, 2005) and Watkins and Makin (1996).

5.5 A RETROSPECTIVE EXAMINATION OF THESE EXPERIMENTAL RESULTS

These experiments provide a basis for several important discussion points. The following commentary uses the advantage of over 30 years of hindsight to interpret the experimental results and possible explanations. My interpretations of the results are also biased by the context of this book and an earlier synopsis in Toole (2006). I will begin with a look at the experimental variable factors.

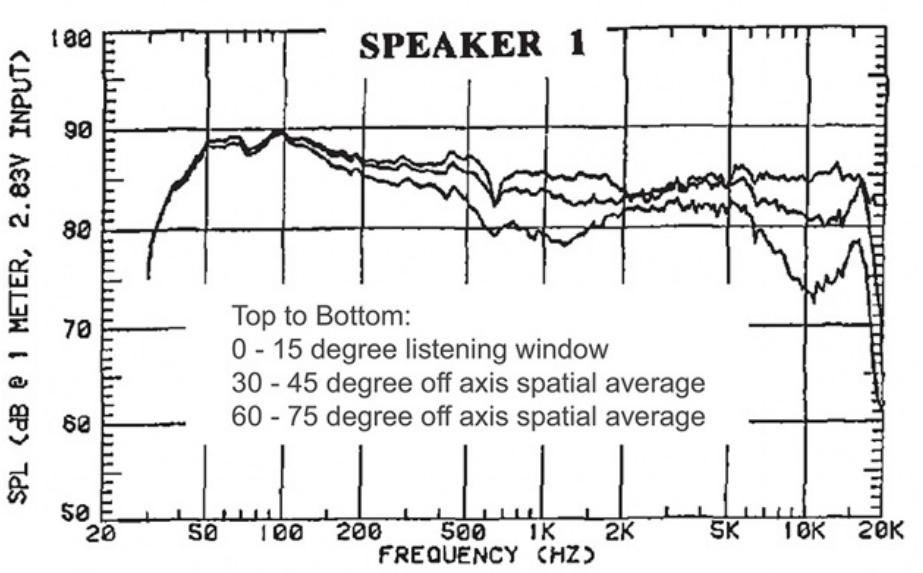
5.5.1 The Tested Loudspeakers

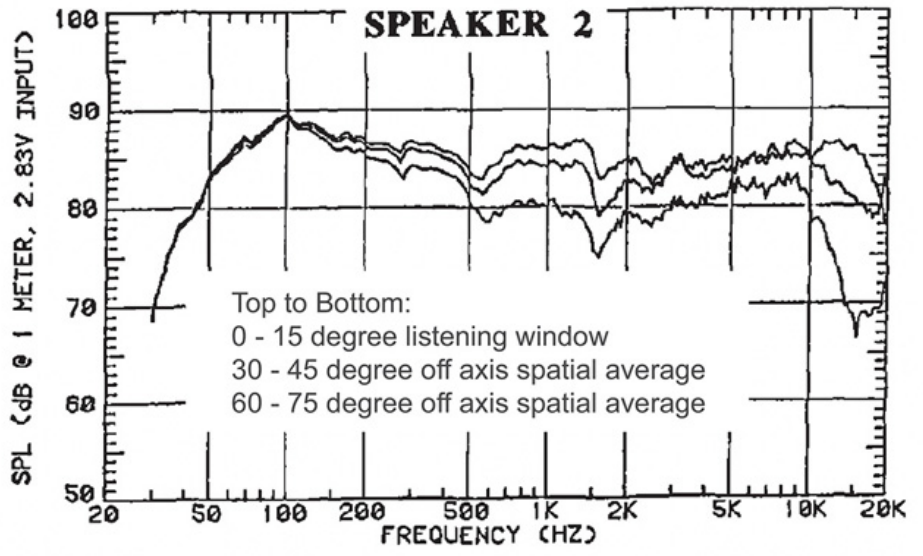
Some basic NRCC format anechoic measurements on the three test loudspeakers are shown in Olive et al. (1995) and are reproduced here in Figure 5.9.

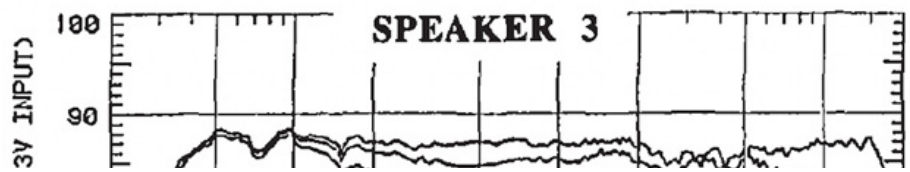
All three loudspeakers are well engineered in the sense that they exhibit no evidence of audible medium- or high-Q resonances. This is a significant fact, because the remaining identifying characteristics are broadband effects.

Loudspeakers 1 and 2 show slightly elevated broadband bass output, which can be attractive to listeners at moderate sound levels (due to equal-loudness contours), but loudspeaker 2 has the lowest output at very low frequencies. Loudspeaker 1 has a sagging off-axis output in the mid frequencies, which accentuates a very low-Q spectral hump between 2 and 6 kHz which is likely to be audible. This is a very common attribute of two-way loudspeakers. In contrast, loudspeaker 2 has a slightly elevated output

in the midrange, a slight “knee” around 1.5 kHz followed by a gently rising high-frequency output, although it is not excessive. Loudspeaker 3 is commendably smooth, tending to flat overall, but with a broadband listening window dip between 2.5 and 5 kHz. Directivity is not shown as a separate measure, but all three share rising mid-frequency directivity, followed by wider dispersion when the small tweeters appear above about 2 kHz.







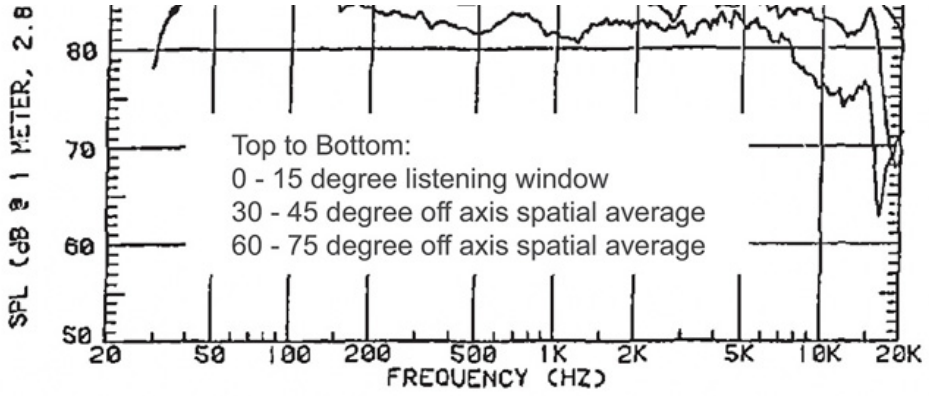


FIGURE 5.9 *Anechoic measurements on the three loudspeakers used in the listening tests. Figure 2 from Olive et al. (1995) with descriptive notes added for clarity.* ⏎

How are these likely to be rated? There is no "bad" loudspeaker here, just three approaches to saleable products. All these loudspeakers will have audible "personalities" but ones that are probably easily adapted to in real life, because annoying resonances are absent. The variations in bass tilt and extension might be differentiating factors.

5.5.2 The Rooms and Loudspeaker/Room Interactions

As can be seen in Figure 5.8, the four rooms employed here are all distinctive. This is a variable with substantial differences.

The acoustical differences among the rooms are revealed in measurements of the loudspeakers in them. Measuring the steady-state room curve at the listening location should shed some light on listening preferences, especially at low frequencies where we know they are most meaningful. Figure 5.10 shows the data. Above the transition frequency region around 400–500 Hz, as we know, two ears and a brain react very differently than an omnidirectional microphone (Figure 1.5), so the details in room curves are not reliable evidence of potential sound quality.

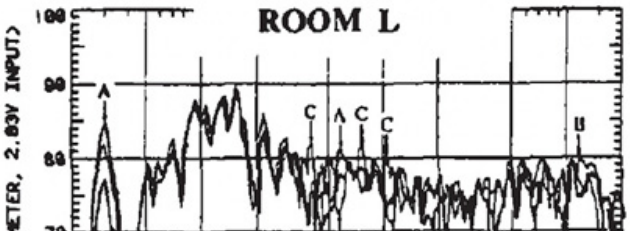
There appears to be a labeling problem with these curves. According to the original caption, curve A is for loudspeaker 2, B is loudspeaker 1, and C is loudspeaker 3. However, in these curves the loudspeaker exhibiting the lowest output at very low frequencies is shown as B, but this means that B

is really loudspeaker 2, as shown in Figure 5.9, and A is probably loudspeaker 1. This now makes sense as the low bass curves for loudspeakers 1 (A) and 3 (C) are similar, and 2 (B) has lower output.

These curves show the expected room resonance roughness at low frequencies, and there are significant differences in the energy in those modes. The most preferred rooms, L and A, have the most bass, and the least preferred, B and IEC, have the least bass, but the amount of bass emphasis that these listeners found to be attractive is interesting—too much bass was judged to be better than too little. Room L is the largest, and is "L" shaped, so it has acoustical complexity that the others don't. The result, for whatever reason, is an elevated upper bass hump from about 80 to 250 Hz and another elevation above about 5 kHz. Neither would seem to be desirable. Room A is very small, and the listener is seated close to the rear wall, with the loudspeakers close to the front wall. The listeners therefore experienced a huge bass boost at the first-order length mode around 40 Hz, and they liked the "one note" bass—bad bass is better than insufficient bass.

Looking at the summarized ratings in Figure 5.11, it is possible to visualize some of the statistical conclusions.

Figure 5.11(a), for the binaural version of Experiment One, in which the multiple loudspeaker comparison tests were completed in one room before moving to another. The conclusion was that "Room" was not a statistically significant factor, as can be seen here in the similar overall heights of the bars for all rooms. Within each room there are changes in loudspeaker preferences, but loudspeaker 1 never achieves the highest rating, whereas loudspeaker 2 does consistently well. Ratings for loudspeaker 3 vary somewhat, winning once, but it comes in as a second-place choice on average. Many differences were not statistically significant as indicated by the 95% confidence error bars.



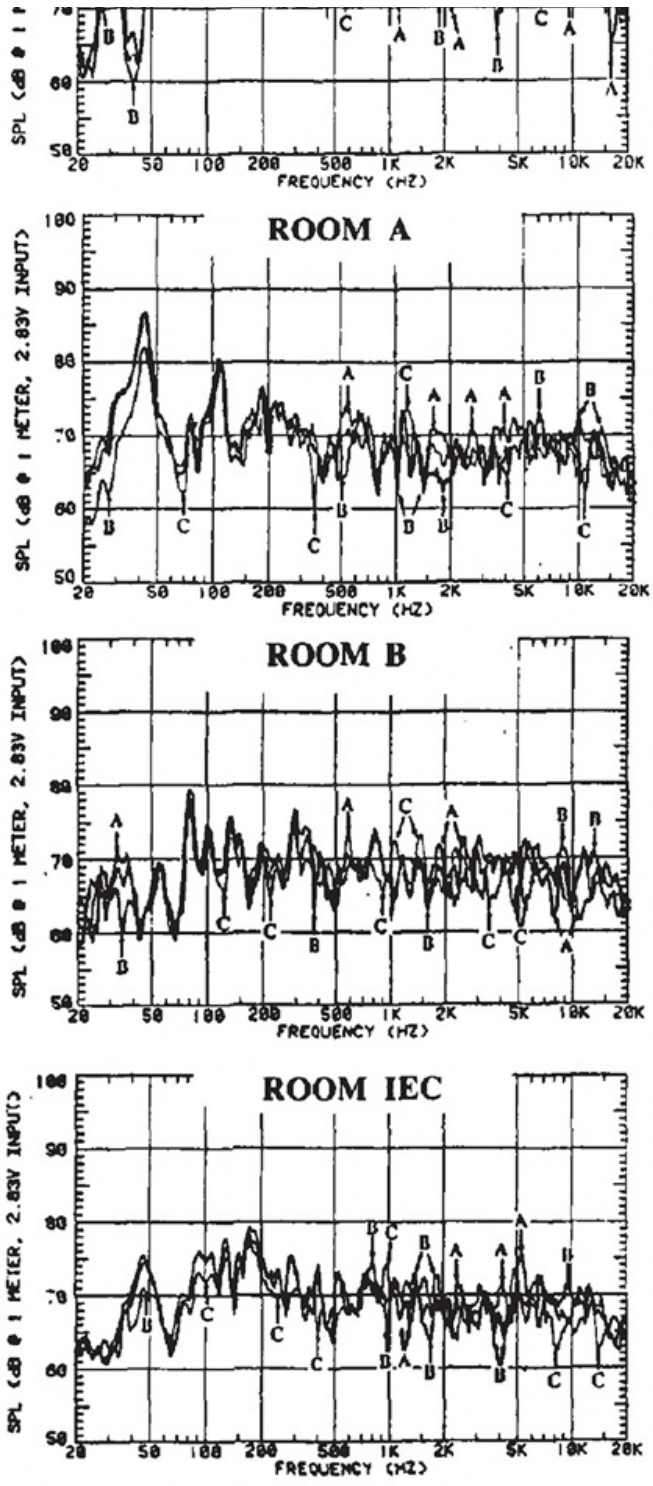


FIGURE 5.10 *The 1/6-octave smoothed steady-state room curves for the three loudspeakers in position 2 (center) in the four rooms. From Olive et al. (1995, figure 19). The room data are*

arranged top to bottom in descending order of how listeners rated the rooms in terms of preference. ⏎

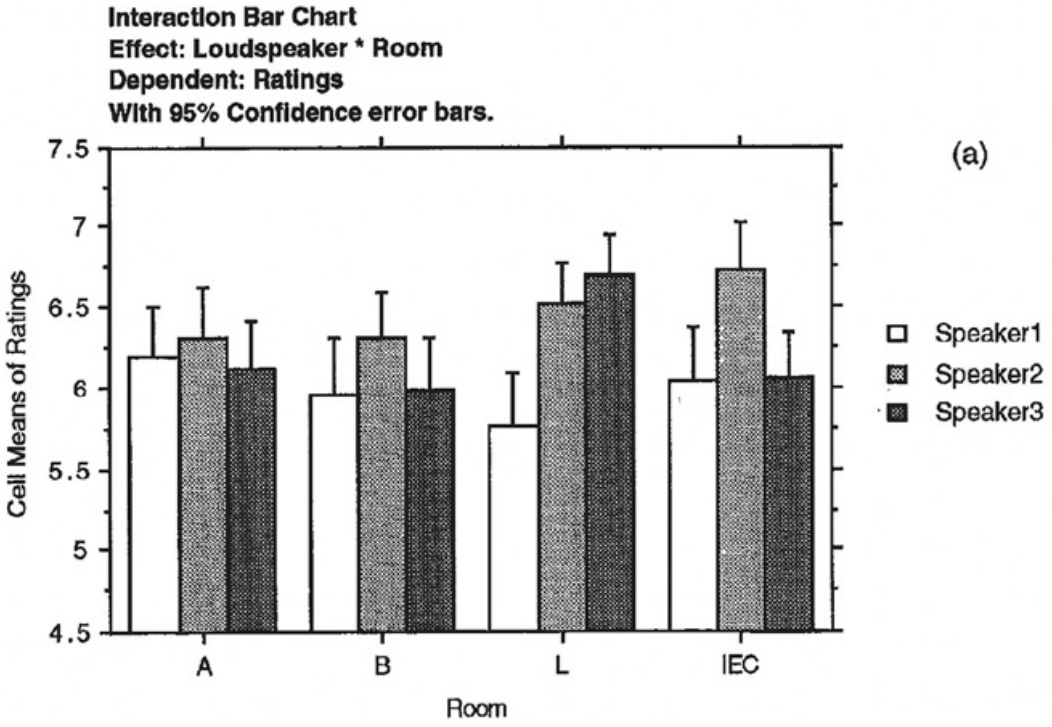
Conclusion: the rooms had an effect on the absolute ratings but did not dominate the relative ratings, the ranking of the loudspeakers; the loudspeakers did. The results of the live version resulted in the same ranking as this binaural version, demonstating the competence with which the experimental method was applied. During the comparison tests, listeners were able to separate the sound of the room from the overall experience to such an extent that it was sufficiently normalized to allow them focus on differences among the loudspeakers. The multiple-loudspeaker experimental method almost certainly assisted in this process, and it is possible that there was some adaptation to, familiarization with, room sound. Seeing the large differences in room performance in Figure 5.10, this is a remarkable feat.

Figure 5.11(b), for binaural Experiment Two, in which the sound of the loudspeaker is combined with that of the room, and different loudspeaker/room combinations were evaluated in rapid succession. There was no opportunity to subjectively normalize the listening room, and the differences in ratings across rooms were large. The variations in rooms were more audible than the differences among loudspeakers. The listeners clearly responded favorably to elevated bass, as rooms L and A were rated highly, and rooms B and IEC—the more "normal" rooms—were not rated highly. When they liked the combined sound, as in A and L (the elevated bass), the differences between loudspeakers all but disappeared. When bass levels were lower, ratings dropped and differences among the loudspeakers were more evident, as in B and IEC. Figure 5.10 shows that the elevated bass that the listeners approved of was not what would be considered to be "smooth"—it was significantly flawed, but it was sufficiently attractive that the listeners substantially ignored other broadband differences among the loudspeakers. They were all rated identically. For many listeners, excessive bass is a "forgivable sin", if not an advantage. When the bass was deficient,

other factors in loudspeaker sound quality became audibly significant. This alone is a thought-provoking result of these experiments.

Interesting and possibly connected is the fact that a gently rising bass output is common in nature and desirable in sound reproducing systems (Toole, 2015). The highest-rated forward-firing loudspeakers exhibit slightly downward tilted steady-state room curves; see Figure 3.15.

BINAURAL 1



BINAURAL 2

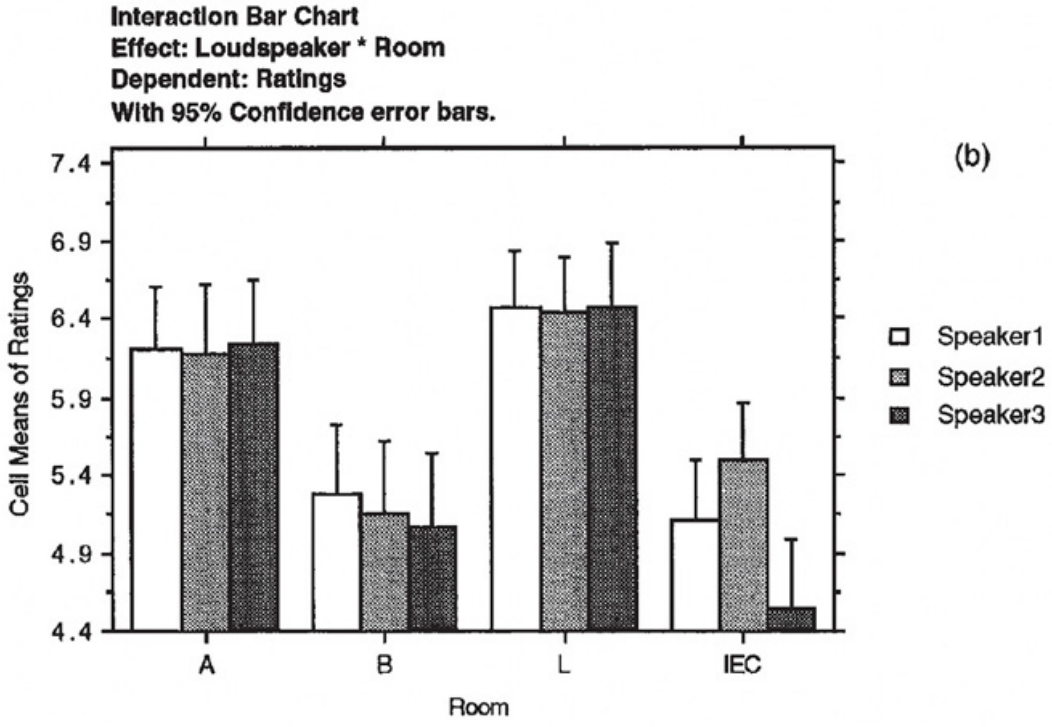


FIGURE 5.11 *Bar graphs showing the mean preference ratings for the three loudspeakers as a function of "Room". (a) shows results of the binaural version of Experiment One, and (b) shows results for binaural Experiment Two.*

It should not be concluded that listeners can always adapt to or normalize rooms to the extent shown here. As will be seen, there is always some adaptation, but in Experiment One it was enabled by the experimental method of comparing different loudspeakers within each room. Experiment Two is more closely related to real-life listening, in that the sounds of the room and the loudspeaker are combined. Listeners living in the room may develop enough familiarity with it to be able to discern some characteristics of loudspeakers being listened to, but having multiple comparison loudspeakers available is a very advantageous situation. The tests had the additional advantage of being in mono.

5.5.3 Program

Program is a major factor because recordings vary in bandwidth, spectral balance, spectral density, loudness fluctuations and so on. These interact with all other variables. If the recording lacks low bass, the advantage of a wide-bandwidth loudspeaker is not revealed, or a problem room resonance is not excited. If a recording has predominantly vocal or single instrument content, troublesome loudspeaker resonances may not be revealed. Spectral balance variations such as excessive or insufficient treble or bass obviously bias the evaluations of loudspeakers, and they interact differently with room acoustic characteristics. There are many such interactions happening in listening tests.

Hearing bass is advantageous, so exaggerated room resonances can benefit a loudspeaker with deficient bass output, even if the combined bass is colored. Some contemporary music styles, studio recordings, demand substantial bass output. Inadequate bass will be criticized and excessive bass rewarded.

In these and all responsible listening tests the programs are chosen to reveal differences, not to disguise them (see Toole and Olive, 1988, and Olive, 1994). Several different recordings are used, averaging the results. Recordings with obvious spectral or bandwidth abnormalities are avoided so large program dependent effects are less likely. Also, loudspeaker comparisons are performed "within" individual programs, so although program is a variable, it is not one that is instantly compared. There will be some normalization of their effects during a comparison sequence. Listeners also get familiar with the program excerpts as they are repeated countless times throughout many listening tests. They become test signals—all pleasure from them is lost.

5.5.4 The Experimental Method

As pointed out back in Section 1.4, the manner of conducting subjective evaluations can influence the results of those evaluations. This was a consideration in these tests, and because of carefully done binaural recordings we have contrasting results from two fundamentally different kinds of subjective evaluations. Schuck et al. (1993) also compared the monophonic multiple-loudspeaker evaluation method with the traditional paired-comparisons method, finding that there were substantial time and labor savings with the multiple-loudspeaker method, an increased ability to discern differences and no statistical disadvantages.

Statistical measures give us confidence in observed differences within variables and their independence. Here, the multiple-loudspeaker method provided a perceptual advantage for listeners, who in one sitting experienced three different loudspeaker "sounds" within a relatively stable room sound, and program selection. What this experiment has shown is that listeners are able to separate the spectrum that is changing (the sounds from the different loudspeakers) from those that are constant (the colorations added by the room within which the evaluations are being done) and the recorded selection being played. In earlier discussions reference has been made to Bregman (1999), who describes this as a form of auditory streaming in what he called "Auditory Scene Analysis".

Of course, different programs will elicit different responses based on spectral interactions, and their abilities to reveal resonance or other problems in the loudspeakers. *But radically changing the room size and shape had no significant effect on the loudspeaker sound quality ratings.* This is profound. But keep in mind that this is a multiple-loudspeaker evaluation exercise, not at-home recreational listening, or single-stimulus subjective reviewing. It demonstrates the power of the method in revealing aspects of sound quality.

In these experiments, loudspeaker position effects required multiple listening sessions with each loudspeaker auditioned in each position, and the results averaged. In each of the rooms, the room modes associated with the length dimension would remain relatively constant during the comparisons, with the width and tangential modes varying as the loudspeakers were randomly moved among the three (lateral) positions during the test. Residual differences would be averaged out to some extent, so even the low-frequency modes, which are a significant factor in normal listening, would be attenuated in their influence in these results.

Making the randomized positional changes caused interruptions in the process, and listeners would lose some of their concentration and adaptation during the intervals. A positional substitution apparatus, in which the active loudspeaker is automatically moved to the same position, is an enormous advantage, saving time and reducing variability in results. This was done at Harman (Olive et al., 1998) and at Samsung, McMullin et al. (2015). There is an ambitious hobbyist version described in an audio forum, MatthewS (2023), so corporate funding is no longer a requirement. Anyone performing frequent subjective evaluations is well advised to make such an investment. Reviewers take note.

5.6 ADAPTATION HAPPENS QUICKLY IN EVERYDAY LIFE

In the real world, speech is always monophonic, a single source somewhere in a room. Reproduced speech is frequently monophonic, coming from a

center channel in a movie or TV broadcast. Small radios and portable audio devices are also functionally monophonic. Speech, considered as a signal, consists of complex modulated resonances in the larynx and vocal tract, combined with plosive and fricative (consonant) sounds. The intelligibility of such a signal requires discerning subtle differences in real time while listening through rooms. This is clearly related to listening to speech, singing and instrumental music through loudspeakers in rooms. Figure 12.9 shows that the directivity of human voices is similar to that of forward-firing cone and dome loudspeakers, confirming the relevance of the following information.

Lochner and Burger (1958), Soulodre et al. (1989) and Bradley et al. (2003) found that multiple room reflections contribute to improved speech intelligibility. The most elaborate of these experiments used an array of eight loudspeakers in an anechoic chamber to simulate early reflections and a reverberant decay for several different rooms (Bradley et al., 2003). The result was that early reflections (>50 ms) had the same desirable effect on speech intelligibility as increasing the level of the direct sound. Remarkably, even attenuating the direct sound had little effect on intelligibility in a sound field with sufficient early reflections.

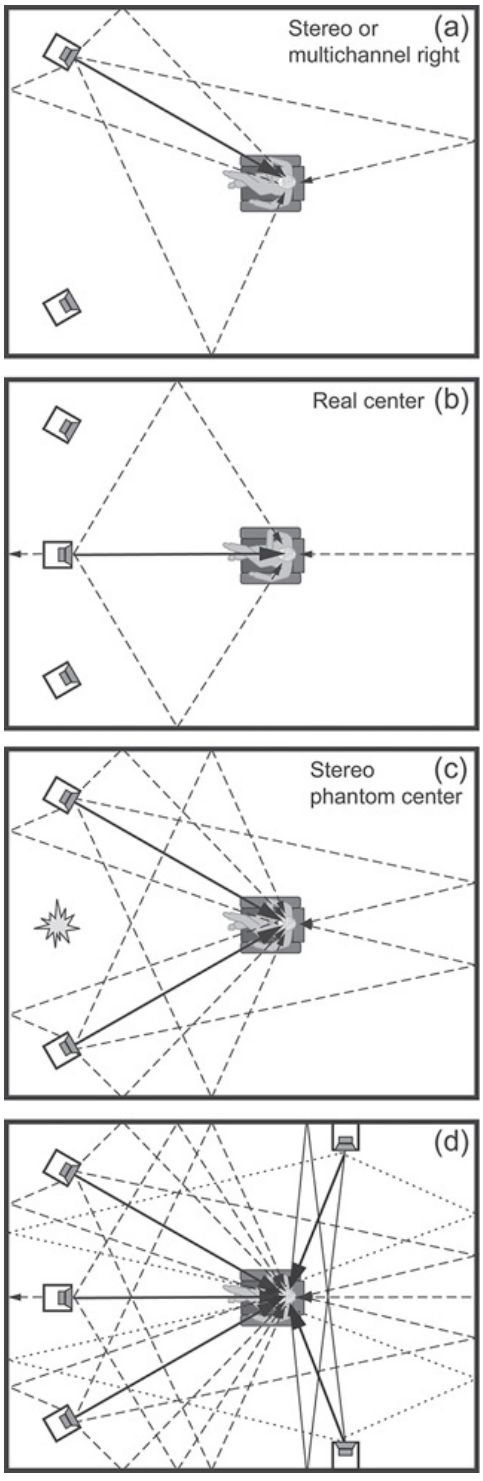
Isolating reverberation time as a factor in school classrooms, it was found that optimum speech communication occurred for RT in the range 0.2 to 0.5 s (Sato and Bradley, 2008). This is conveniently in the range of normally furnished domestic rooms and recording control rooms.

Brandewie and Zahorik (2010) found that speech intelligibility improved by about 18% when listeners were given prior exposure to reverberant rooms, leading them to conclude that "the physical effects of acoustic reflections may be suppressed via high-level perceptual processes that require adaptation time to the particular reflective listening environment in order to be effective." In subsequent experiments even more demanding tests confirmed this result, noting that "the processes underlying this effect are relatively rapid—on the order of seconds—and do not show longer term improvement" (Srinivasan and Zahorik, 2012; Zahorik and Brandewie, 2016).

These investigations point to the conclusion that normal listening rooms and control rooms do not impair speech intelligibility, rather they are more likely to improve it—at least when they originate in simple (monophonic) sources as commonly happens. Familiarity with the room matters—adaptation to small rooms is real, effective and quick.

5.7 MORE CHANNELS, MORE PLEASURE AND MORE CONFUSION

Increasing the number of channels expands the soundstage. Stereo gives us a frontal spread of sound images, with impressions of space and a hint of envelopment. Ear-level multichannel reproduction expands that to include sides and rears, and a generous sense of space and persuasive envelopment. Immersive multichannel expands the soundstage into the height dimension, envelopment can be quite realistic, and immersion very engaging. These are positive attributes, but every level of added sound field complexity both rewards and challenges the binaural auditory system in ways that might not appear obvious.



Direct sound and early reflections:

(a) For hard-panned right sound image or a monophonic loudspeaker.

Listeners are very critical of loudspeaker sound quality.

(b) For hard-panned center sound image or a monophonic loudspeaker.

Listeners are very critical of loudspeaker sound quality.

(c) For stereo phantom center image and to varying degrees for other phantom images across the soundstage.

Listeners are less critical of loudspeaker sound quality when listening to conventional recordings and both channels are active.

(d) For a 5-channel multichannel system, ***all channels active.***

Listeners are much less critical of loudspeaker sound quality when listening to conventional recordings and all channels are active.

FIGURE 5.12 *Direct sound and first reflection sound ray diagrams for stereo system and a 5-channel surround-sound home theater—with all channels actively radiating sound. The reality is even more complicated because there are floor and ceiling*

reflections associated with the direct sounds shown by bold lines. The comments in italics refer to results of subjective evaluations discussed in Sections 4.4 and 4.5. ⏎

Figure 5.12 illustrates the direct sounds and first reflections for the loudspeakers in stereo and conventional 5-channel home theaters. These are the second-loudest sounds listeners will hear from each loudspeaker, arriving later in time and from different angles. In general, for forward-firing loudspeakers, sounds radiated to the rear are mainly low frequencies. For this reason, they are a less significant factor, but for bipole, dipole, and omnidirectional designs they will obviously be more influential.

Apparent in this illustration is the dramatic simplification in the reflection pattern of a real center channel loudspeaker (b) compared to the "double-mono" phantom center image (c). They cannot sound the same in terms of timbre or space, not even considering the audible interaural crosstalk dip at 2 kHz discussed in Section 4.3.2.

In the complex acoustics of many listening rooms, it is easy to imagine that the binaural hearing system might be confused at localizing a phantom image where there is no physical sound source that matches the collection of reflected sounds. For most soundstage images there is also the perceptual conflict caused by direct-sound HRTF cues related to sounds arriving from the wrong directions: ±30°. The hearing system copes with these anomalies amazingly well, as experience attests. However, it cannot be easy, compared to the room interaction "purity" of a single (mono) center source, as in (b), or a hard-panned left or right location (a). In fact, it is the perceptual simplicity of a center loudspeaker that creates conflict when it is combined with double-mono phantom images at other locations.

As mentioned earlier, this presents a dilemma for recording engineers doing multichannel mixes: to use a real center channel or a phantom center image. According to King et al. (2023), there is no simple solution. It is difficult to make a real center image fit into a soundstage consisting of phantom images. It seems that we have adapted to the distortions inherent in stereo phantom image reproduction to such an extent that when we hear

an uncompromised image in the same soundstage it stands out. It is often not preferred to the familiar double-mono phantom center.

One finds evidence of this in several multichannel music recordings, some of which include 5-channel setup signals, but when listening to the program nothing comes from the center loudspeaker. When viewing music concert videos with multichannel soundtracks, put your ears close to the center loudspeaker. Sometimes it is silent, other times the featured artist emerges from all three front loudspeakers—evidence of an unresolved problem. Still, for the simplistic popular music format of center plus hard-panned L and R combined with ambiance, a real center should be an improvement.

It is frequently assumed by recording engineers that center channel loudspeakers in home systems are inferior to the L and R fronts, which is sometimes true. This is an unfortunate situation, a consequence of poor marketing among other things. Selling multichannel to people with stereo systems was a challenge, and it was thought it would be easier if the additional loudspeakers could be small and inexpensive. This ignores the reality that in movies the center channel is arguably the most important channel, delivering the bulk of dialog and much of the on-screen sound, while anchoring the soundstage for off-center listeners. In cinemas, all front (screen) loudspeakers are identical in performance, as they are in well-designed home theaters.

5.8 MULTIPLE “ROOMS” MODIFY PERCEPTIONS

In Sections 4.4 and 4.5 it was shown that from the point of view of sound quality, the most challenging listening tests for loudspeakers were those done in mono. People heard and complained about resonances in loudspeakers. Adding a single channel for stereo made listeners less responsive to flaws in loudspeakers, as persuasively shown in Section 4.4. Adding more channels in multichannel made them even more forgiving (Figure 5.7(b)). Simple logic suggests that the complexity of the sound field in the room was responsible. The ray diagrams in Figure 5.12 are

persuasive evidence of that. But, if the resonances were in all the loudspeakers, why would having more of them playing simultaneously make the resonances less audible?

Toole and Olive (1988) compared resonance detection thresholds in both stereo and mono in a normally reflective listening room using a compound loudspeaker source that could vary the direct-to-reflected sound ratio over a large range. The differences in direct and reflected sound ratios (i.e. the complexity of the sound field) in the various modes of listening was enormous. But no effect was seen in the results. Detection thresholds for resonances were the same in stereo and mono, and with large variations in direct-to-reflected ratio.

This contradicts other tests reported here (Section 4.4) that showed an increased tolerance for resonances in loudspeakers in stereo tests.

A possible explanation is that in the Toole and Olive (1988) experiments the test signals were technical sounds, pink noise and pulses, not commercial recordings. With technical signals both loudspeakers radiate the same sounds, and room reflections repeat the resonance cues in those sounds, providing multiple "looks" for the perceptual process to respond to. These repetitions contributed to the ability to hear resonances, and monophonic and stereo listening were judged to be similarly revealing of those resonances.

However, in stereo *music* recordings the loudspeakers do not radiate the same sounds. A stereo recording has monophonic components: hard-panned left and right channels, and a double-mono center channel and other phantom images. Added to these sounds are real or synthesized reflections to deliver the perception of large venues. These superimposed acoustical space cues have no relationship to the listening room, so the normal association that listeners have to the listening room is altered. The listening room is fixed, but the "rooms" in recordings and movies change, and are typically louder than the natural room reflections. Delivering persuasive impressions of being enveloped in larger spaces is important in music and in many scenes in movies.

Figure 5.13 is an attempt to describe this process in a functional block diagram, inspired by discussions with co-author Todd Welti.

These are limited experimental data, and it is risky to extrapolate too far on this basis alone. However, this result suggests that stereo itself, the added channel, is not responsible for the degraded ability to detect resonant flaws in loudspeakers. Instead, it is the distinctive features in stereo recordings, where the added channels and loudspeakers bring spatial dimensions that are not present in monophonic listening. These sounds are in the strong, direct sounds from loudspeakers in stereo and multichannel systems, and in listening room reflections of those sounds. The direct and room-reflected sounds all contain resonance cues associated with the loudspeakers, so resonances in the loudspeakers are there to be heard, but our perception of them appears to be inhibited. More channels deliver more persuasive large-room spatial impressions. Olive and Martens (2007) found no evidence of adaptation in multichannel evaluations that used commercial recordings.

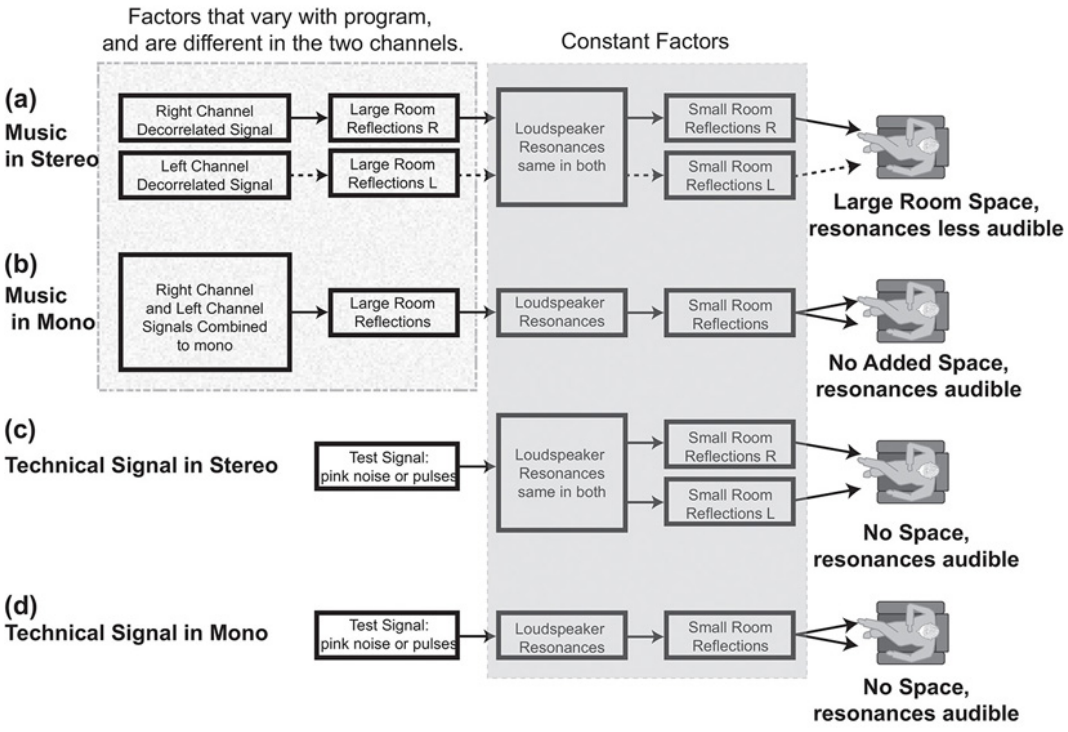


FIGURE 5.13 *A block diagram showing the factors involved in forming subjective judgments of sound quality in different experimental circumstances. Figures 5.13(a) and (b) describe the two music-based methods, one in stereo and one in mono. In stereo there are differences in the sounds in the two channels that provide the listener with a left-to-right soundstage accompanied by recorded reflections and reverberation consistent with the impression of large acoustical spaces. The sounds from the two channels are significantly decorrelated when they reach the listener's ears, resulting in impressions of large space envelopment. In mono, the left and right channel signals are combined to mono, and the combination is delivered through a single loudspeaker. The listener's ears receive the same, highly correlated, sounds, and the only binaurally perceived sense of space is that of the small listening room. Figures (c) and (d) show the comparable situation when the test signal is pink noise or pulses. These are monophonic sources, delivered through one or two loudspeakers, and in both cases the only binaurally perceived sense of space is that of the listening room. Across all of these situations, the one that reduced the listener's ability to detect resonances is (a), stereo music delivered through two loudspeakers.* ⏎

Two observations may explain the disparity.

1. Perceptually, we are more influenced by the "large room" spatial information in recordings than by those in the small listening room (Hughes et al., 2016). Therefore, it seems plausible that this results in us diverting our attention from the reflections in the listening room to the strong and temporally distinctive large-space recorded reflections.
2. Chapter 7 explores the importance of space in our perceptions, and there is much evidence showing that we prefer recorded "large-room"

sounds at higher levels and longer delays than small room reflections can provide. These spatial impressions can be delivered by stereo and even more persuasively by multichannel systems.

The summary statement of Section 7.1.3 is:

> In summary, acoustically "normal" listening rooms appear to be good starting points. Listeners preferred much higher levels of reflections than are provided by listening rooms, supporting a need for multichannel delivery systems. Preferences indicated in these tests indicate that 30 to 50% of overall satisfaction is contributed by sound quality and 50 to 70% by non-specific "feelings of space".

When we listen to systems and recordings capable of delivering convincing impressions of large space envelopment, are our perceptional priorities modified? Are we less aware of timbral subtleties? In practice it really does not matter because we only listen to commercial recordings. The issue is settled—adding more channels results in us being more forgiving of loudspeaker flaws. The exception is when there are intervals of monophonic sound, which happen when sounds are hard-panned to individual loudspeakers. All stereo and multichannel music and movies include periods of strongly, if not completely, monophonic sound. These intervals provide opportunities for listeners to respond to early reflection patterns associated with the relevant loudspeaker in the listening room. There is evidence in the experiments of Section 4.4 that this happens, meaning that the evaluation of loudspeakers should always be performed in the most revealing format: mono. Loudspeakers highly rated in mono listening continue to be highly rated as channels are added, but the reverse may not be true (see Figure 4.13).

5.8.1 Speculation

All this happened in small listening rooms. Does it happen in large multichannel venues like cinemas? Is the sound field of a concert hall

sufficiently complex to compromise one's ability to discern fine timbral features in live voices and instruments?

Benade, in Figure 1.6, acknowledges the remarkable ability of humans to recognize qualities in the sound source while listening through the complex sound field of a room, even a concert hall. This is undoubtedly true, but it may also be true that sharing the experience among hundreds of listeners in a large reverberant venue compromises what qualities can be discerned. This unorthodox thought needs consideration.

Any normalization of early reflection patterns in small listening rooms is fundamentally different from normalization that might take place with 100 musicians distributed over a large stage in a concert hall. Large halls deliver first reflections from boundaries with delays an order of magnitude longer than those in small rooms. Only the reflection from the stage floor is "early" in the small room sense. Reflections are encouraged in concert halls, reverberation times are relatively high and the sound field significantly diffuse. The challenge for listeners to normalize the room sound is considerable.

We know that relatively close microphone locations and skilled mixing by recording engineers can deliver rewarding loudspeaker listening experiences in small rooms. They are certainly different, but not necessarily worse in all respects. In fact, if one looks at air attenuation alone (see Figure 12.19) the combined high-frequency losses in the direct and reflected sounds in large concert halls means that much high-frequency detail is missing in the live experience for most of the audience. Headphone listening is even more different. In both cases one can hear details that are inaudible in most of the performance environment.

Concert halls have been used for several "live vs. reproduced" loudspeaker demonstrations starting over a century ago, and all were judged to have been successes at the time (see Sections 4.2 and 4.3). Listening in concert venues flattered loudspeakers that would be seriously criticized when reproducing recordings in domestic spaces.

Is it possible that listening in large spaces that deliver persuasive envelopment causes us to become less able to hear and evaluate subtleties

in sound quality? Possibly so, but the live experience is uniquely satisfying. Envelopment is magnificent, bass is smooth and deep, and the visual and social experiences are not trivial.

5.8.2 Implications for Car Audio

An automobile cabin is a distinctive space for listening. The programs are generally in stereo, but there is no symmetrical stereo seat, so nobody hears uncompromised stereo. Surround presentations are advantageous and upmixing is common. Listeners are in the acoustical near field of some of the loudspeaker transducers (with their very adjacent boundaries), in a space that has some strongly reflective windows, but which is overall acoustically well damped and filled with large sound scattering seat backs and headrests. The bass in cars can be excellent as there are no very low-frequency cabin modes, although they and mechanical resonances exist at higher frequencies. The transducers are located for convenience within the structure of the automobile, not for an optimal integrated performance as separate full-bandwidth channels. A lot of signal processing is required to achieve anything resembling the auditory experience one has in a normal room. Perhaps this is why premium car audio systems are increasingly using immersive systems to deliver engaging experiences.

It is possible to conclude that, in car audio, multichannel delivery systems, incorporating upmixing, are advantageous to the overall listening experience—maximizing directional and spatial interest while reducing the audibility of resonant colorations in transducers and automobile structures. It may be sufficient to achieve a reasonable spectral balance, including adequate low bass, and most customers will be pleased.

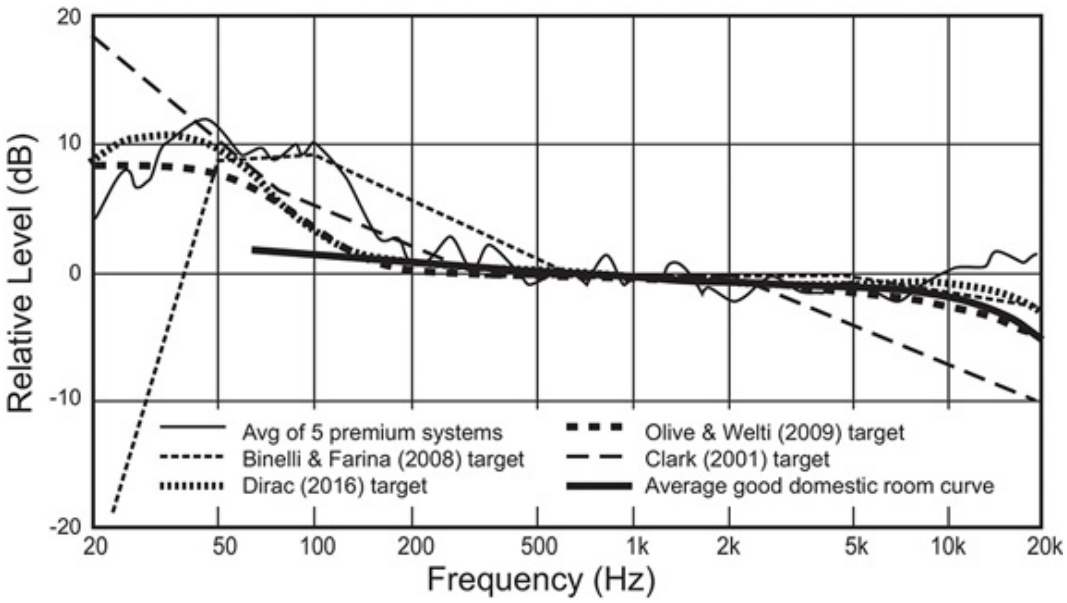


FIGURE 5.14 *Four in-cabin steady-state target curves used in the design of car audio systems, compared to some real-world examples, and to an average room curve for highly rated domestic loudspeakers.* ⏎

Figure 5.14 shows a few of many possible target curves for automobile audio systems. All show the effect that in-car background noise has on the target frequency responses. Figure 9.8 shows that the low-frequency background noise from a vehicle in motion exceeds the average spectrum of foreground listening to popular music below about 200 Hz, therefore requiring a boost to maintain audibility of the bass frequencies. In a stationary car the bass is likely to be perceived as excessive, so some vehicles incorporate a speed- or noise-related bass adjustment to automatically maintain a reasonable balance.

These observations will likely change as car manufacturers explore variations on binaural sound delivery using loudspeakers in the headrests. This reduces weight and allows for independent program choices for occupants. There are new dimensions of perceptual rewards and problems.

5.8.3 Implications for Headphone Listening

Headphone listening is a distinctive experience. Most program material is recorded for stereo loudspeaker reproduction, not headphones, so the directional, spatial and distance impressions, while entertaining, are not what were intended. What is experienced is called *lateralization*, in-head localization, as opposed to normal sound *localization* which occurs for externalized sound images. It is what I researched for my PhD thesis (Sayers and Toole, 1964; Toole and Sayers, 1965a, 1965b). The simple interaural amplitude and time differences (IAD and ITD) determining image locations are understood. It is very different from sound localization in the three-dimensional sound field of rooms, where there is the perceptually essential precedence effect (Section 6.1) in which the first sound arrival from a source, the direct sound, determines where it is localized. The direct sound, not the steady-state room curve, also dominates the perception of sound quality (Sections 2.4 and 11.3). To experience these direct-sound phenomena through headphones the recordings must be head-tracked binaural (dummy head) recordings capable of reconstructing the perception of a room sound field. The fact that these listening room–related spatial effects are absent in headphone listening to stereo recordings makes it a substantially different perceptual experience, complicating attempts to find measurable correlations between loudspeaker and headphone listening.

Without a listening room, the only room impressions are those in the recordings themselves, which is beneficial, but they are not delivered as intended, through loudspeakers. The timbral distortions caused by loudspeaker phantom image acoustical crosstalk are absent (Section 4.3.2), so sound quality is improved. Room resonances are absent, a further advantage, but the physical sensations of bass are missing, which is a disadvantage.

Toole and Olive (1988) found that headphone and anechoic listening yielded the highest detection thresholds for resonances, meaning that listeners are relatively forgiving of headphone resonances. Fortunately, in-ear headphones are generally free of resonances over most of the frequency range. On- and around-ear designs have more complications because of the acoustically complex pinnae. Most designs are susceptible to variations in

bass level due to air leaks in the interface with variable human anatomies and measurement fixtures (Toole, 1984 and Chapter 15). Such broadband spectral trends in frequency response are easily heard and are likely to be judged against long-term listening experience and auditory memory, with program being a factor. Finding the optimum headphone target is work in progress, but just as in loudspeakers, programs are not uniform so access to tone controls is a useful feature. Sean Olive will address this topic in detail in Chapter 15.

5.8.4 Implications for Low-Frequency Room Resonances

Deep, clean, tight bass is universally desirable but very often prevented by the small rooms in which we listen—all of which are different in structure and in arrangements of loudspeakers and listeners. Section 5.1 explains the distinctive sound fields that exist above and below the transition/Schroeder frequency. The earliest investigations revealed a positive correlation between low-bass extension and overall sound-quality ratings (Figure 2.5). More recent investigations showed that bass performance accounted for approximately 30% of overall sound-quality ratings, and both bass extension and quality were significant factors (Olive, 2004a, 2004b). The degree to which listeners can adapt to irregular, resonant bass performance has been investigated by a few studies, and it seems that while some amount of adaptation is possible, there are limits. Todd Welti will explain the details in Chapter 14.

5.9 CONCLUSION

There is much to think about but it is clear that enjoying music and movies involves perceptual trade-offs. Some audiophiles strive to reproduce the original concert hall performance while the masses enjoy the music as delivered, many never hearing the “real thing”.

Music, speech, drama and special effects all have survived generations of technically inferior loudspeakers, small rooms and headphones. Humans are very adaptable creatures, but why compromise if it is not necessary?

REFERENCES

Baskind, A. and Polack, J-D. (2000). "Sound Power Radiated by Sources in Diffuse Field", *108th Convention, Audio Eng. Soc*., Preprint 5146.

Binelli, M. and Farina, A. (2008). "Digital Equalization of Automotive Sound Systems Employing Spectral Smoothed FIR Filters", *Audio Eng. Soc. Convention*, Paper 7575.

Bradley, J. S., Sato, H. and Picard, M. (2003). "On the Importance of Early Reflections for Speech in Rooms", *J. Acoust. Soc. Am*., vol. 113, pp. 3233–3244. http://dx.doi.org/10.1121/1.1570439

Brandewie, E. and Zahorik, P. (2010). "Prior Listening in Rooms Improves Speech Intelligibility", *J. Acoust. Soc. Am*., vol. 128, pp. 291–299.

Bregman, A. S. (1999). *Auditory Scene Analysis, the Perceptual Organization of Sound*, MIT Press

Büchlein, R. (1962). "The Audibility of Frequency Response Irregularities", reprinted in English in *J. Audio Eng. Soc*., vol. 29, pp. 126–131 (1981).

Clark, D. (2001). "Progress in Perceptual Transfer Function Measurement —Tonal Balance", *Audio Eng. Soc. Convention*, Paper 5407.

Fryer, P. A. (1975). "Intermodulation Distortion Listening Tests", *J. Audio Eng. Soc*. (abstracts), vol. 23, p. 402.

Fryer, P. A. (1977). "Loudspeaker Distortions, Can We Hear Them", *Hi-Fi News Rec. Rev*., vol. 22, pp. 51–56.

Hughes, R., Cox, T., Shirley, B. and Power, P. (2016). "The Room-in-room Effect and Its Influence on Perceived Room Size in Spatial Audio Reproduction", *141st Convention Audio Eng. Soc*., Preprint 9621.

ITU-R BS.775-3 (2012). "Multichannel Stereophonic Sound System with and without Accompanying Picture", *International Telecommunication Union*, Geneva.

ITU-R BS.1116-3 (2015). "Methods for the Subjective Assessment of Small Impairments in Audio Systems", *International Telecommunication Union*, Geneva.

King, R., Theriault, M. and Massenburg, G. (2023). “A Practical Approach to the Use of Center Channel in Immersive Music Production”, *155th Convention Audio Eng. Soc*., Paper 168. ⏎

Lochner, J. P. A. and Burger, J. F. (1958). “The Subjective Masking of Short Time Delayed Echoes by Their Primary Sounds and Their Contribution to the Intelligibility of Speech”, *Acustica*, vol. 8, pp. 1–10. ⏎

MatthewS (2023). “Blind Listening Test 2”. www.audiosciencereview.com/forum/index.php?threads/blind-listening-test-2-neumann-kh-80-vs-jbl-305p-mkii-vs-edifier-r1280t-vs-rcf-arya-pro5.43343/. Accessed August 2024. ⏎

McMullin, E., Celestinos, A. and Devantier, A. (2015). “Environments for Evaluation: The Development of Two New Rooms for Subjective Evaluation”, *139th Convention, Audio Eng. Soc*., Preprint 9460. ⏎

Olive, S. E. (1994). “A Method for Training Listeners and Selecting Program Material for Listening Tests”, *Audio Eng. Soc. 97th Convention*, Preprint 3893. ⏎

Olive, S. E. (2004a). “A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 1—Listening Test Results”, *116th Convention, Audio Eng. Soc*., Preprint 6113. ⏎

Olive, S. E. (2004b). “A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 2—Development of the Model”, *117th Convention, Audio Eng. Soc*., Preprint 6190. ⏎

Olive, S. E., Castro, B. and Toole, F. E. (1998). “A New Laboratory for Evaluating Multichannel Audio Components and Systems”, *105th Convention, Audio Eng. Soc*., Preprint 4842. ⏎

Olive, S. E., Hess, S. M. and Devantier, A. (2008). “Comparison of Loudspeaker-Room Equalization Preference for Multichannel, Stereo, and Mono Reproductions: Are Listeners More Discriminating in Mono?”, *Audio Eng. Soc. 124th Convention*, Preprint 7492. ⏎

Olive, S. E. and Martens, W. L. (2007). “Interaction between Loudspeakers and Room Acoustics Influences Loudspeaker Preferences in

Multichannel Audio Reproduction", *123rd International Convention, Audio Eng. Soc.*, Preprint 7196. Also Olive's PhD dissertation, McGill University, Montreal, Canada, 2007. ⏎

Olive, S. E., Schuck, P. L., Sally, S. L. and Bonneville, M. (1992). "The Effects of Loudspeaker Placement on Listeners' Preference Ratings", *93rd Convention, Audio Eng. Soc.*, Preprint 3352. ⏎

Olive, S. E., Schuck, P. L., Sally, S. L. and Bonneville, M. (1994). "The Effects of Loudspeaker Placement on Listeners' Preference Ratings", *J. Audio Eng. Soc.*, vol. 42, no. 9, pp. 651–668. ⏎

Olive, S. E., Schuck, P. L., Sally, S. L. and Bonneville, M. (1995). "The Variability of Loudspeaker Sound Quality Among Four Domestic Sized Rooms", *99th Convention, Audio Eng. Soc.*, Preprint 4092. ⏎

Olive, S. E. and Welti, T. (2009). "Validation of a Binaural Car Scanning System for Subjective Evaluation of Automotive Audio Systems", *Audio Eng. Soc., 36th International Conference: Automotive Audio.*

Pike, C., Brookes, T. and Mason, R. (2013). "Auditory Adaptation to Loudspeaker and Listening Room Acoustics", *Audio Eng. Soc., 135th Convention*, Preprint 8971. ⏎

Sato, H. and Bradley, J. (2008). "Evaluation of Acoustical Conditions for Speech Communication in Working Elementary School Classrooms", *J. Acoust. Soc. Am.*, vol. 123, pp. 2064–2077. ⏎

Sayers, B. McA. and Toole, F. E. (1964). "Acoustical Image Lateralization Judgments with Binaural Transients", *J. Acoust. Soc. Amer.*, vol. 36, pp. 1199–1205. ⏎

Schroeder, M. R. (1954). "Statistical Parameters of the Frequency Response Curves of Large Rooms", *Acustica*, vol. 4, pp. 594–600. Translated from the German original in *J. Audio Eng. Soc.*, vol. 35, pp. 299–306 (1987). ⏎

Schroeder, M. R. (1996). "Schroeder Frequency Revisited", *J. Acoust. Soc. Am.*, vol. 99, pp. 3240–3241. ⏎

Schuck, P., Olive, S., Toole, F., Sally, S., Bonneville, M., Momtahan, K. and Verrault, E. (1993). "Perception of Perceived Sound in Rooms: Some

Results of the Athena Project", *12th International Conference, Audio Eng. Soc*., Paper 12-006. ⏎

Soulodre, G. A., Popplewell, N. and Bradley, J. S. (1989). "Combined Effects of Early Reflections and Background Noise on Speech Intelligibility", *J. Sound Vib*., vol. 135, pp. 123–133. ⏎

Srinivasan, N. K. and Zahorik, P. (2012). "Prior Listening Exposure to a Reverberant Room Improves Open-set Intelligibility of High-variability Sentences", *J. Acoust. Soc. Am*., vol. 133, pp. EL33–EL39 (JASA Express Letters). ⏎

Toole, F. E. (1984). "The Acoustics and Psychoacoustics of Headphones", *Audio Eng. Soc. 2nd International Conference*, Paper C1006. ⏎

Toole, F. E. (1985). "Subjective Measurements of Loudspeaker Sound Quality and Listener Preferences", *J. Audio Eng. Soc*., vol. 33. pp. 2–31. ⏎

Toole, F. E. (1986). "Loudspeaker Measurements and Their Relationship to Listener Preferences", *J. Audio Eng. Soc*., vol. 34, pt. 1, pp. 227–235, pt. 2, pp. 323–348. ⏎

Toole, F. E. (2006). "Loudspeakers and Rooms for Sound Reproduction—A Scientific Review", *J. Audio Eng. Soc*., vol. 54, pp. 451− 476. ⏎

Toole, F. E. (2015). "The Measurement and Calibration of Sound Reproducing Systems", *J. Audio Eng. Soc*., vol. 63, pp. 512–541. This is an open-access paper available to non-members at www.aes.org www.aes.org/e-lib/browse.cfm?elib=17839 ⏎

Toole, F. E. and Olive, S. E. (1988). "The Modification of Timbre by Resonances: Perception and Measurement", *J. Audio Eng. Soc*., vol. 36, pp. 122–142. ⏎

Toole, F. E. and Sayers, B. McA. (1965a). "Laterization Judgments and the Nature of Binaural Acoustic Images", *J. Acoust. Soc. Am*., vol. 37, pp. 319–324. ⏎

Toole, F. E. and Sayers, B. McA. (1965b). "Inferences of Neural Activity Associated with Binaural Acoustic Images", *J. Acoust. Soc. Am*., vol. 37, pp. 769–779. ⏎

Watkins, A. J. (1991). “Central, Auditory Mechanism of Perceptual Compensation for Spectral-envelope Distortion”, *J. Acoust. Soc. Am.*, vol. 90, pp. 2942–2955. ⏎

Watkins, A. J. (1999). “The Influence of Early Reflections on the Identification and Lateralization of Vowels”, *J. Acoust. Soc. Am.*, vol. 106, pp. 2933–2944. ⏎

Watkins, A. J. (2005). “Perceptual Compensation for Effects of Echo and of Reverberation on Speech Identification”, *Acta Acustica United with Acustica*, vol. 91, pp. 892–901. ⏎

Watkins, A. J. and Makin, S. J. (1996). “Some Effects of Filtered Contexts on the Perception of Vowels and Fricatives”, *J. Acoust. Soc. Am.*, vol. 99, pp. 588–594. ⏎

Zahorik, P. and Brandewie, E. (2016). “Speech Intelligibility in Rooms: Effect of Prior Listening Exposure Interacts with Room Acoustics”, *J. Acoust. Soc. Am.*, vol. 140, pp. 74–86. ⏎

CHAPTE
R 6

The Perception and Measurement of Reflections

Floyd Toole

DOI: 10.4324/9781003477495-7

6.1 A SINGLE REFLECTION: THE PRECEDENCE EFFECT AND MORE

Investigations into our remarkable ability to localize sources of sound in acoustically hostile (i.e. reflective) environments have been recorded more than a century ago. In audio, the names “Haas effect” and the “law of the first wavefront” have been used to identify this phenomenon, but current scientific work has settled on the other original term, precedence effect. Whatever it is called, it describes the well-known phenomenon wherein the first arrived sound, normally the direct sound from a source, dominates our impression of where sound is coming from.

Within a time interval often called the “fusion zone”, later arriving sounds are suppressed in terms of their contributions to perceived direction,

there is only one perceived sound image, although its character may be modified. The later arrivals add to overall loudness and contribute to sound quality. They are not masked.

Sounds arriving after the fusion interval may be perceived as spatially separated auditory images, coexisting with the direct sound, but the direct sound is still perceptually dominant. At very long delays the secondary images are perceived as echoes, separated in time as well as direction. The literature is not consistent in language, with the word "echo" often being used to describe a delayed sound that is not perceived as being separate in either direction or time.

I just searched "precedence effect" on the internet, and there are some misunderstandings about it, as well as a few explanations that reference our research, even showing figures from earlier editions of this book, without attribution. Misunderstandings are inevitable because the early experiments —and some later ones—simplistically employed identical sounds and sound levels for the direct and later arriving sound (the reflection). In reality, natural reflections are always lower in sound level than the direct sound and do not always have the same frequency spectrum, depending on the directional characteristics of the sound source and the acoustical nature of the reflecting surface. The early experiments also used speech, a logical sound, but not representative of the spectrally and temporally distinctive sounds in music and movies. More tests were needed. The full explanation of the precedence effect is multidimensional. A thorough academic survey is in Brown et al. (2015), which goes far beyond what is needed for our purposes. Let us begin with the Haas experiments that popularized the precedence effect.

Haas was not the first person to observe the primacy of the first arrived sound so far as localization in rooms is concerned (Gardner, 1968, 1969, 1973), but work done for his PhD thesis in 1949, translated from German to English in Haas (1972), has become one of the standard references in the audio field. Let us review his core experiment.

Figure 6.1 shows the essence of the experiment. On the hemi-anechoic space provided by the flat roof of a laboratory building, a listener faced

loudspeakers placed 45° apart. The Haas (1972) translation describes the setup as “at an angle of 45° to the left and right side of the observer” (p. 150). This could be construed in two ways. Gardner (1968) in a translation of a different Haas document reports “loudspeakers … at an angle of 45°, half to the right and half to the left of him”. When Lochner and Burger (1958) repeated the Haas experiment they used loudspeakers separated by 90°. So, there is ambiguity about the angular separation. The arrangement shown here seemed like the most logical one. Because the experiments were done on a flat roof, to minimize the effect of the roof reflection Haas placed the loudspeakers directly on the roof, aimed upwards towards the listener’s ears. He found, though, that there was no significant difference if the loudspeakers were elevated to ear level, and that is the configuration used for the experiments. The experiments therefore included one reflection from each loudspeaker, the floor bounce.

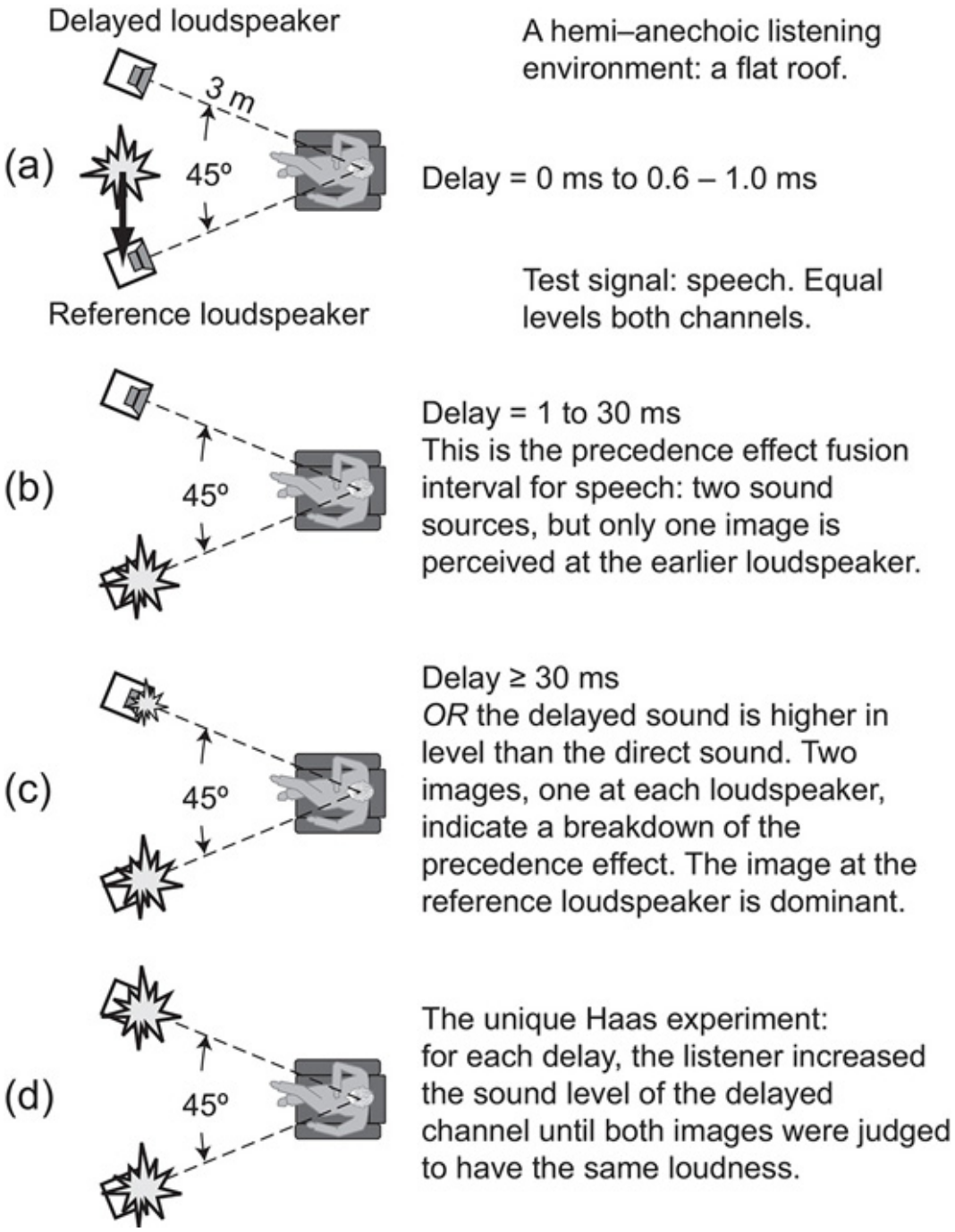


FIGURE 6.1 *A progression of localization effects observed in the experimental setup used by Haas, including stereo (summing) localization, the precedence effect and the equal-loudness experiment.* ⏎

A recording of running speech was sent to both loudspeakers, and a delay could be introduced into the signal fed to one of them. In all situations except that for Figure 6.1(d), both signals were radiated with the same sound level, which, of course, cannot happen in real life, but this is a paradigm repeated in several academic investigations. It was not a real-world test, but a a good thesis project in 1949.

Figure 6.1(a): summing localization. When there is no delay, the perceived result was a phantom (stereo) image floating midway between the loudspeakers. When delay was introduced, the center image moved towards the loudspeaker radiating the earlier sound, reaching that location at delays of about 0.7–1.0 ms. This is called "summing localization" and it is the basis for the phantom images that can be positioned between the left and right loudspeakers in stereo recordings, assuming a listener in the symmetrical listening location (Blauert, 1996). As discussed in Section 4.3.1 the panning of sound images in stereo recordings is most often done using interchannel amplitude differences.

Figure 6.1(b): For delays in excess of 1 ms, it is found that the single image remains at the reference loudspeaker up to about 30 ms. *This is the precedence effect*: when there are two (or more) sound sources and only one sound image is perceived. It needs to be noted that the 30 ms interval is only for speech, and only for identical equal level direct and delayed sounds, as determined in hemi-anechoic space. In reality, reflected sounds are always lower in sound level, which lengthens the fusion zone. The fusion zone also varies with the type of sound, being significantly shorter for transient sounds, and longer for sounds including reverberation or auditioned in a reflective space, as will be shown later. In spite of these limitations, many people think that 30 ms is a definitive number, and wrongly use it indiscriminately.

Figure 6.1(c): multiple images—the breakdown of precedence. With delays greater than 30 ms, but certainly by 40 ms, the listener becomes aware of a second sound image at the location of the delayed loudspeaker. The precedence effect has broken down because there are two images, but

the second image is a subordinate one; the dominant (louder) localization cue still comes from the loudspeaker radiating the earlier sound.

Figure 6.1(d): the Haas equal-loudness experiment. In all illustrations to this point the first and delayed sounds had identical amplitudes. But Haas moved even farther from passive acoustical realities and deliberately amplified the delayed sound, as would happen in a professional audio sound reinforcement situation. It is not a facsimile of any possible single reflection. His interest was to determine how much higher in sound level the delayed sound could be before it became the dominant localization cue (i.e. subjectively louder). To do this he asked his listeners to adjust the sound level of the delayed loudspeaker until both of the perceived images appeared to be equally loud. This is the balance point, beyond which the delayed loudspeaker would be perceived as being dominant. The objective was to prevent an audience from seeing a person speaking in one direction, and being distracted by a louder voice coming from a sound reinforcement loudspeaker in a different direction.

Figure 6.2 shows that from about 5 ms to about 25 ms the later loudspeaker can be as much as 10 dB higher in level before it is perceived to be equally loud, and therefore a major distraction to the audience. Naturally, this would depend on where the audience member is seated relative to the symmetrical axis of the two sound sources.

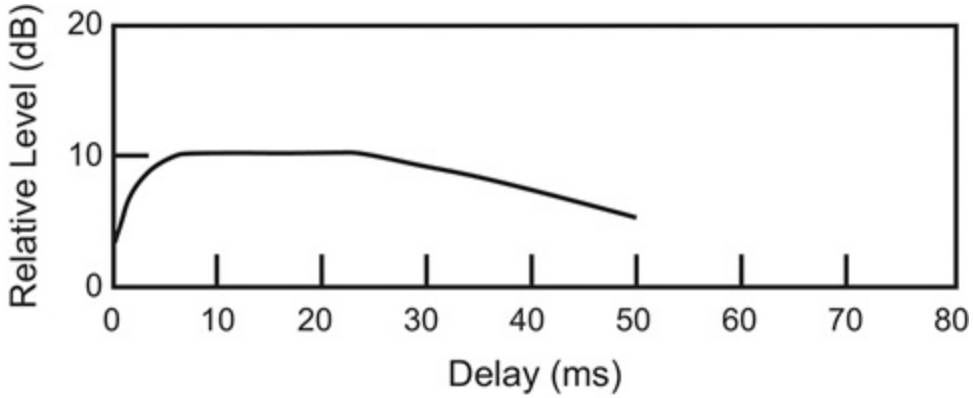


FIGURE 6.2 *The sound level of the delayed sound, relative to that of the first arrival, at which listeners judged the two sound images to be equal in loudness.* ⏎

Haas described this as an “echo suppression effect”. Some people have mistakenly taken this to mean that the delayed sound is masked—it isn’t. Within the precedence effect fusion interval, there is no masking—all of the reflected (delayed) sounds are audible, making their contributions to timbre and loudness, but the early reflections simply are not heard as spatially separate events. They are perceived as coming from the direction of the first sound—this, and only this, is the essence of the “fusion”. The widely held belief that there is a “Haas fusion zone” within which everything gets innocently combined, is simply untrue.

Haas noted that there were audible effects having nothing to do with localization. First, the addition of a second sound source increased loudness. There were some changes to sound quality “liveliness” and “body” (Haas, 1972, p. 150), and a “pleasant broadening of the primary sound source” (Haas, 1972, p. 159). Increased loudness was a benefit to speech reinforcement, and the other effects would be of concern only if they affected intelligibility.

Benade (1985) contributed a thoughtful summary under the title “Generalized Precedence Effect”, stating:

1. The human auditory system combines the information contained in a set of reduplicated sound sequences and hears them as though they were a single entity, provided (a) that these sequences are reasonably similar in their spectral and temporal patterns and (b) that most of them arrive within a time interval of about 40 ms following the arrival of the first member of the set.
2. The singly perceived composite entity represents the accumulated information about the acoustical features shared by the set of signals (tone color, articulation, etc.). It is heard as though all the later arrivals were piled upon the first one without any delay, that is, the perceived time of arrival of the entire set is the physical instant at which the earliest member arrived.

3. The loudness of the perceived sound is augmented above that of the first arrival by the accumulated contributions from the later arrivals. This is true even in the case when one of more of the later signals is stronger than the first one to arrive, that is, a strong later pulse does not start a new sequence of its own.
4. The apparent position of the source of the composite sound coincides with the position of the source of the first-arriving member of the set, regardless of the physical directions from which the later arrivals may be coming.
5. If there are any arrivals of sounds from the original acceptably similar set which come in after a delay of 100–200 ms, they will not be accepted for processing with their fellows. On the contrary, they will be taken as a source of confusion and will damage the clarity and certainty of the previously established percept. These "middle-delay" signals that dog the footsteps of their betters may or may not be heard as separate events.
6. If for some reason a reasonably strong member of the original set should come in with a delay of something more than 250–300 ms, it will be distinctly heard as a separate echo. This late reflection will be so heard even if it is superposed on a welter of other (for example, reverberant) sounds.

It is important to notice that these very strongly worded categorical statements all emphasize that there is an accumulation of information from the various members of the sequence. It is quite incorrect to assume that the precedence effect is some sort of masking phenomenon which, by blocking out the later arrivals of the signal, prevents the auditory system from being confused. Quite to the contrary, those arrivals that come in within a reasonable time after the first one actively contribute to our knowledge of the source. Furthermore members of the set that are delayed somewhat too long actually disrupt and confuse our perceptions even when they may not

be consciously recognized. If the arrivals are later yet, they are heard as separate events (echoes) and are treated as a nuisance. In neither case are the late arrivals masked out.

6.1.1 One Reflection—Several Audible Effects

Figure 6.3 illustrates a "begin at the beginning" experiment, in which the number of variables is minimized. The listening environment was anechoic, not hemi-anechoic with a floor reflection as in the Haas experiments. The signal was English speech, and only a single lateral reflection was examined. The arrangement was intended to be realistic, with the direct sound arriving from the front and the delayed sound arriving from the direction of a wall reflection.

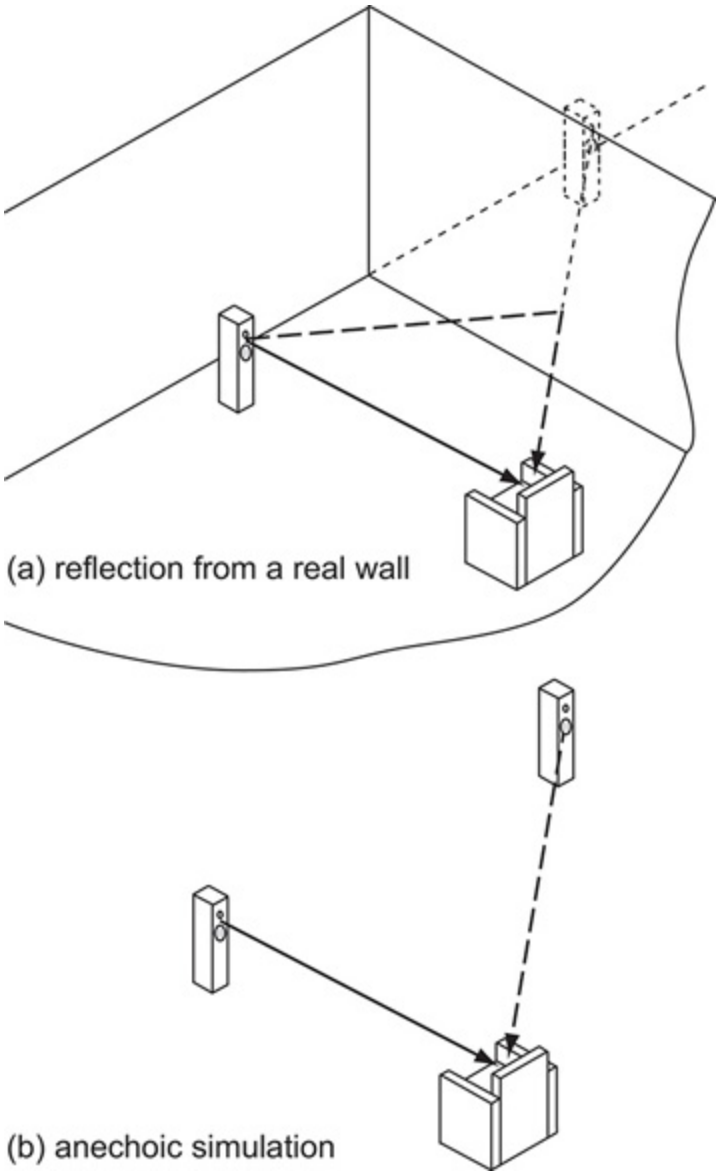


FIGURE 6.3 *An explanation of how an anechoic simulation can imitate a reflection from a real—flat and perfectly reflective—wall. The anechoic setup uses a real loudspeaker to simulate the "mirror image" loudspeaker in the room situation. This is the experimental method that has been used in numerous experiments starting decades ago. Electrical adjustments of delay, amplitude and frequency response of the signal sent to the "reflection" loudspeaker allow the simulation of different geometries and reflective surface types.*

In Figure 6.4 the lowest curve describes the sound level at which listeners reported hearing any change attributable to the presence of the reflection. This is the "absolute threshold"—nothing is perceived for reflections at lower levels. Most listeners described what they heard as a sense of spaciousness (Olive and Toole, 1989). Although the experiment was conducted in an anechoic chamber, a single detectable reflection was sufficient to create the impression of a rudimentary space. Throughout, listeners reported all of the sound as originating at the location of the loudspeaker reproducing the first sound, meaning that the precedence effect was working.

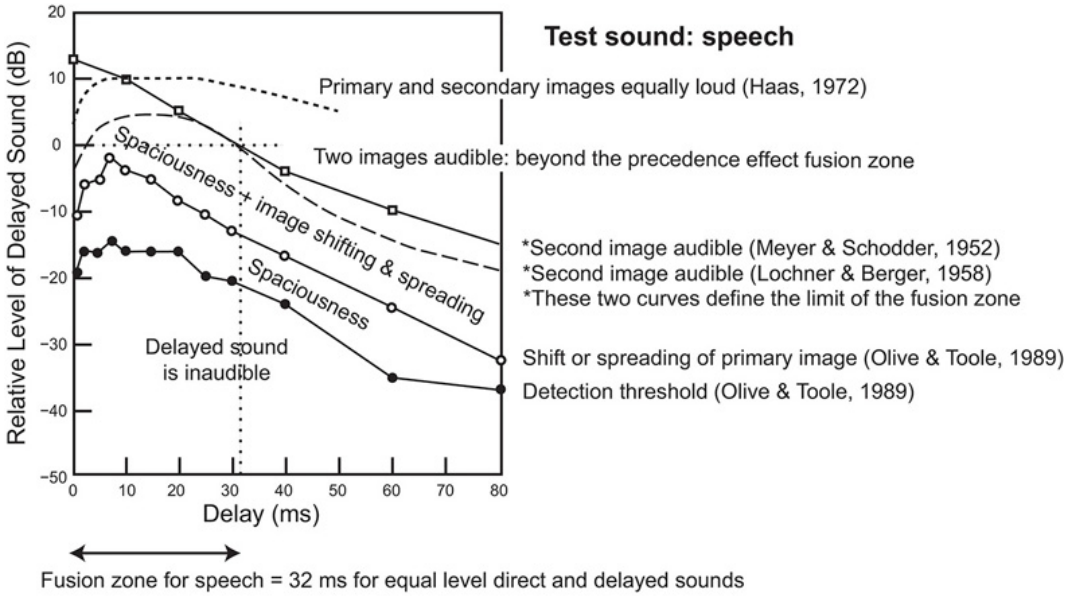


FIGURE 6.4 *An illustration of the several audible effects occurring when a single lateral reflection is added to a direct sound at different times and sound levels, in an anechoic simulation similar to that shown in Figure 6.3(b). All of these curves were determined using speech as a signal. In the experiments, at each of several delays, the sound level of the reflected sound was adjusted to identify those levels at which each of the described perceptions became apparent. The bottom two curves are from Olive and Toole (1989) in which the direct sound was at 0° and the lateral reflection arrived from 65°. Meyer and*

Schodder (1952) had their reflection arrive from 90° and listeners reported when the echo was not perceived at all. Lochner and Burger (1958) employed a direct sound arriving from −45° and a delayed sound from +45° and their listeners reported when the second source was just audible. Adapted from Toole (1990) with additional information from Cremer and Müller (1982, figure 1.25). ⏎

As the sound level of the delayed sound was increased, the impression of spaciousness increased.

The next higher curve is the level at which listeners reported hearing a change in size or position of the main sound image, which the precedence effect causes to be localized at the position of loudspeaker reproducing the earlier sound. This was called the "image shift" threshold. In general these changes were subtle, noticeable in these controlled A vs. B comparisons, but it is doubtful that they would be detected in the context of a multiple-image music or movie soundstage.

As the sound level of the delayed sound was further increased, the impression of spaciousness was further increased and the image modification progressed.

With the two curves that portray the third perceptual category a major transition is reached, because it is at this sound level that listeners report hearing a second sound source or image, simultaneously coexisting with the original one (we have not reached the long delays at which there is a sense of a temporally as well as spatially separate echo) (Lochner and Burger, 1958; Meyer and Schodder, 1952). This means that the precedence-effect directional "fusion" has broken down. Although the original source remains the perceptually louder, directionally dominant source, there is a problem because two spatial events are perceived when there should be only one. It is a reassuring validation of this process that the fusion zone for equal-level direct and delayed speech sounds deduced from these data is about 32 ms, which is I suspect within experimental error of the 30 ms that Haas found (see Figure 6.1(c)).

What can also be seen is that in normal rooms where naturally reflected sounds cannot exceed the level of the direct sound, 0 dB, there are no possible early reflections that would cause the precedence effect to break down. Obviously, such disruptive sounds can occur in multichannel audio recordings—sometimes deliberately.

The top curve is from the well-known work by Haas (1972) in which he asked his listeners to adjust the relative levels of the spatially separate images associated with the direct and reflected sounds until they appeared to be equally loud. This tells us that, in a sound-reinforced public address situation, it is possible to raise the level of delayed sound from a laterally positioned loudspeaker by as much as 10 dB above the direct sound before it is perceived as being as loud as the direct sound. It is important information in the context of professional audio, including sound mixing and processing, but it is irrelevant in the context of small-room acoustics.

All of the data points are thresholds, the sound levels at which listeners detected a change in their perceptions. As will be seen later, some of the perceived changes are beneficial and that, up to a point, listeners find that levels well above threshold provide greater pleasure. For example, the perception described at threshold as "image shift or spreading" may seem like a negative attribute, but when translated into what is heard in rooms it becomes "image broadening" or apparent source width (ASW), highly favored qualities. Even "second image" thresholds can be exceeded with certain kinds of sounds, expanding the size of an orchestra beyond its visible extent in a concert hall, or extending the stereo soundstage beyond the spread of the loudspeakers. In reproduced sound the situation is more confused because there are techniques in the recording process that can achieve similar perceptions. Because all of these factors are influenced by how the recordings are made as well as how they are reproduced, these comments are observations, not judgments of relative merit. Some evidence suggests that even these small effects might be diminished by experience during listening within a given room (Shinn-Cunningham, 2003), another in the growing list of perceptual phenomena that we humans can adapt to.

6.1.2 A Real-World View of the Precedence Effect

If we extract from Figure 6.4 those things that are relevant to sound reproduction from a single loudspeaker, we end up with Figure 6.5. The Haas data have been removed because amplified delayed sounds do not exist in passive acoustics (they can exist in multichannel recordings, though). The "second image" data (Lochner and Burger, 1958; Meyer and Schodder, 1952) have been combined into a single average curve for simplicity. There is some justification for doing this as one curve expresses a "just audible" criterion and the other a "just not audible" criterion.

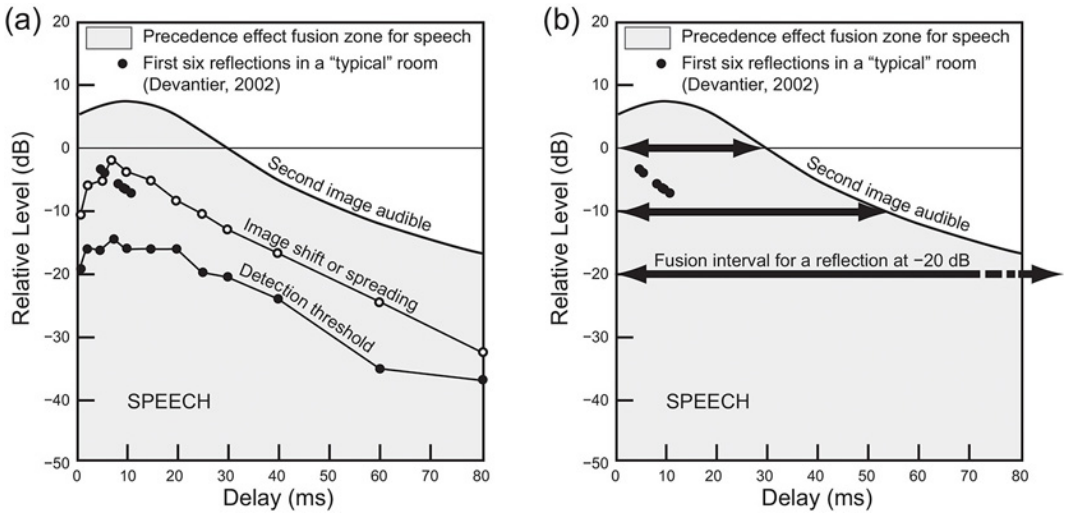


FIGURE 6.5 *(a) is a simplification of Figure 6.4 in which only data relevant to sound in small rooms is preserved, and a shaded area below the second-image curve, representing the precedence-effect fusion zone for speech is identified. This is the range of amplitudes and delays within which a reflected sound will not be identified as a separately localizable event. From Toole (2006). The set of black dots show the delays and amplitudes for the first six reflections in a typical listening room (Devantier, 2002). All are comfortably within the shaded area, indicating that in such rooms the precedence effect is solidly in control of the localization of speech sounds.* ⏎

The area under the “second-image” curve has been shaded. This is the real-world precedence-effect fusion zone for speech, within which any delayed sound will not be perceived as a spatially separate localizable event. This perspective is very different from most discussions of the precedence-effect fusion interval. Normally, only a single number is stated, and that number normally relates to direct and delayed sounds at the same sound levels. This is a correct description of the results of a laboratory experiment, but simply wrong as guidance about what may happen in the real world.

Figure 6.5(b) shows the precedence-effect fusion intervals for delayed sounds at three sound levels. The classic experiments much quoted from psychoacoustic literature generally used equal level direct and delayed sounds—this is the highest large arrow at 0 dB, showing an interval of about 32 ms. In rooms delayed sounds are attenuated by propagation loss, typically −6 dB/double distance, plus sound absorption at the reflecting surfaces. As the delayed sound is reduced in level the fusion zone increases rapidly, to almost 60 ms at a level of −10 dB, and close to 100 ms at a level of −20 dB.

Summarizing, it can be seen in Figure 6.5(b) the much-quoted 30 ms ± as the fusion interval for speech, applies only if the delayed sound has the same level as the direct sound—and let us not forget that the experiments were done using a single reflection in an anechoic space. This is how the early psychoacoustic experiments were conducted, but these circumstances do not represent the acoustical realities in normal rooms. For reflections at realistic sound levels the fusion interval is much longer. In small rooms, the precedence effect is undoubtedly the dominant factor in the localization of speech.

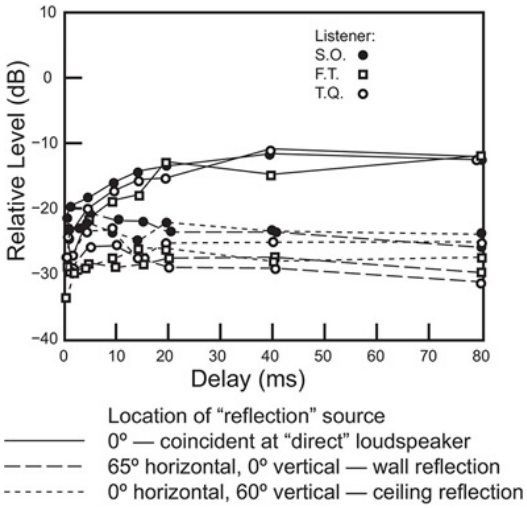
Also of interest is the fact that, in Figure 6.5(a) the typical room reflections are close to, but below, the threshold of detection for any change in size or position of the primary sound image. There is more evidence of this kind in Figure 6.12.

6.1.3 Reflections From Different Directions

Figure 6.6 shows data from Olive and Toole (1989) in which it is seen that the thresholds for the side wall and the ceiling reflections are almost identical. This is counterintuitive, because one would expect a lateral reflection to be much more strongly identified by the binaural discrimination mechanism because of the large signal differences at the two ears. For sounds differing only in elevation, we have only the spectral cues provided by the external ears and the torso (HRTFs).

Figure 6.6 shows that the detection thresholds for reflections from above and from the side are very similar around a relatively constant −25 dB re: the direct sound, with the wall reflection thresholds possibly slightly lower by 2 to 3 dB. In contrast thresholds for the delayed sounds originating in the loudspeaker itself were higher by 10 dB or more. Although the threshold levels themselves might be surprising, intuition is rewarded in that the dominant audible effect of the lateral reflection was spaciousness (the result of interaural differences), and that of the vertical reflection was timbre change (the result of spectral differences). The broadband pink noise used in these tests would be very good at revealing colorations, especially those associated with HRTF differences at high frequencies. On the other hand, continuous noise lacks the strong temporal patterns of some other sounds, like speech.

Detection thresholds for three listeners, using pink noise as a signal, in anechoic chamber simulations.



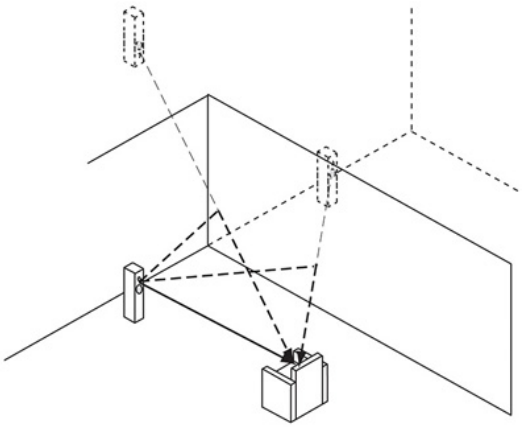


FIGURE 6.6 *The detection thresholds for delayed sounds simulating a wall reflection, a ceiling reflection and one arriving from the same direction as the direct sound. The test signal was pink noise. Adapted from Olive and Toole (1989).* ⏎

Another surprise in Figure 6.6 is that delayed sounds originating in the same loudspeaker are more difficult to hear—the threshold here is consistently higher than for sounds arriving from the side or above, slightly for short delays, and much higher (10+ dB) at long delays. Burgtorf (1961) agrees, finding thresholds for coincident delayed sounds to be 5–10 dB higher than those separated by 40°–80°. Seraphim (1961), used a delayed source that was positioned just above the direct-sound source (~5° elevation difference) and found that, with speech, the threshold was elevated by about 5 dB compared to one at a 30° horizontal separation. The relative insensitivity to coincident sounds appears to be real, and the explanation seems to be that it is the result of spectral similarities between the direct sound and the delayed sound. These sounds take on progressively greater timbral differences as they are elevated (or, one assumes, lowered) relative to the direct sound. For those readers who have been wondering about the phenomenon of "comb filtering", it is worthy of note that this evidence tells us that the situation of maximum annoyance from comb filtering, when the direct and delayed sounds emerge from the same loudspeaker (as happens when the comb filtering is in the program itself), is the one for which we are least sensitive (Clark, 1983).

All this said, it still seems remarkable that a vertically displaced reflection, with no apparent binaural (between the ears) differences, can be detected as well as a reflection arriving from the side, generating large binaural differences. Not only are the auditory effects at threshold different —timbre vs. spaciousness—the perceptual mechanisms required for their detection are also different.

6.2 A REFLECTION IN THE PRESENCE OF OTHER REFLECTIONS

Working with a single reflection allows for intensely analytical investigations but, inevitably, the tests must include others to be realistic. Olive and Toole (1989) examined the audibility of a single lateral reflection simulated in an anechoic chamber with 3 ft (1 m) wedges. For the second experiment, the same physical arrangement was replicated in a typical small room in which the first wall, floor and ceiling reflections had been attenuated using 2-inch (5 cm) panels of rigid fiberglass board. A third experiment was conducted in the same room with most of the absorption removed (mid-frequency reverberation time = 0.4 s). The idea was to show the effects, on the perception of a single reflection, of increasing levels of natural reflected sound within a real room. Figure 6.7 shows the results.

The large changes in the level of reflected sound had only a modest (1 to 5 dB) effect on the absolute threshold or the image-shift threshold of an additional lateral reflection occurring within about 30 ms of the direct sound. At longer delays the thresholds were elevated by up to about 20 dB, a clear response to stronger late-reflected sounds in the increasingly live rooms.

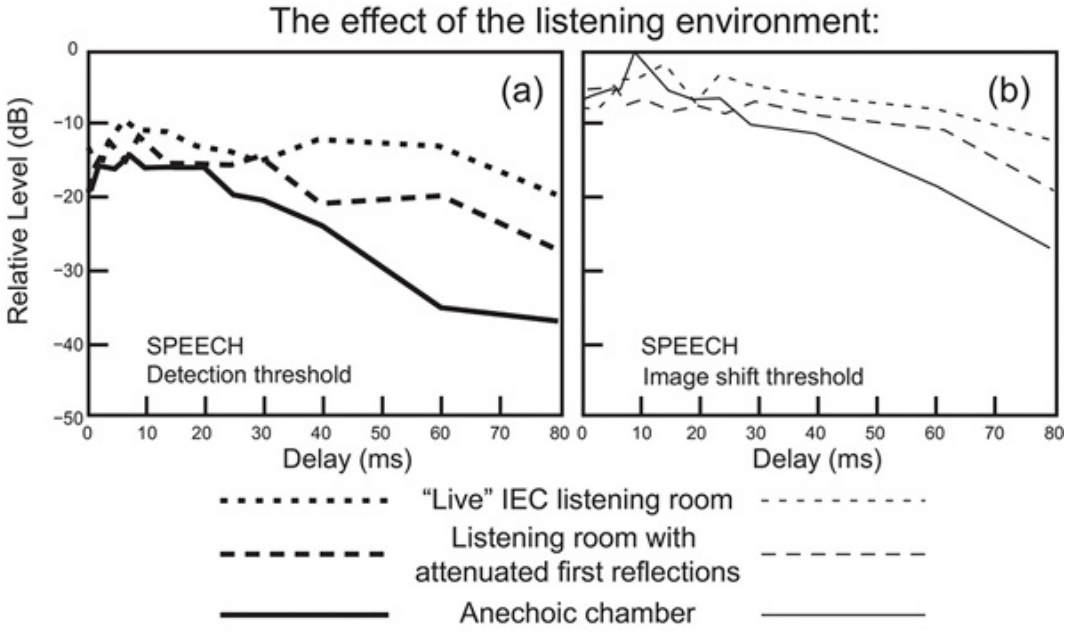


FIGURE 6.7 *Detection and image-shift thresholds as a function of delay for a single reflection auditioned in three very different acoustical circumstances: anechoic, a normal room in which the first-order reflections were attenuated with 2-inch (50 mm) fiberglass board, and the same room in a relatively reverberant configuration (mid-frequency reverberation time = 0.4 s). Olive and Toole (1989).* ⏎

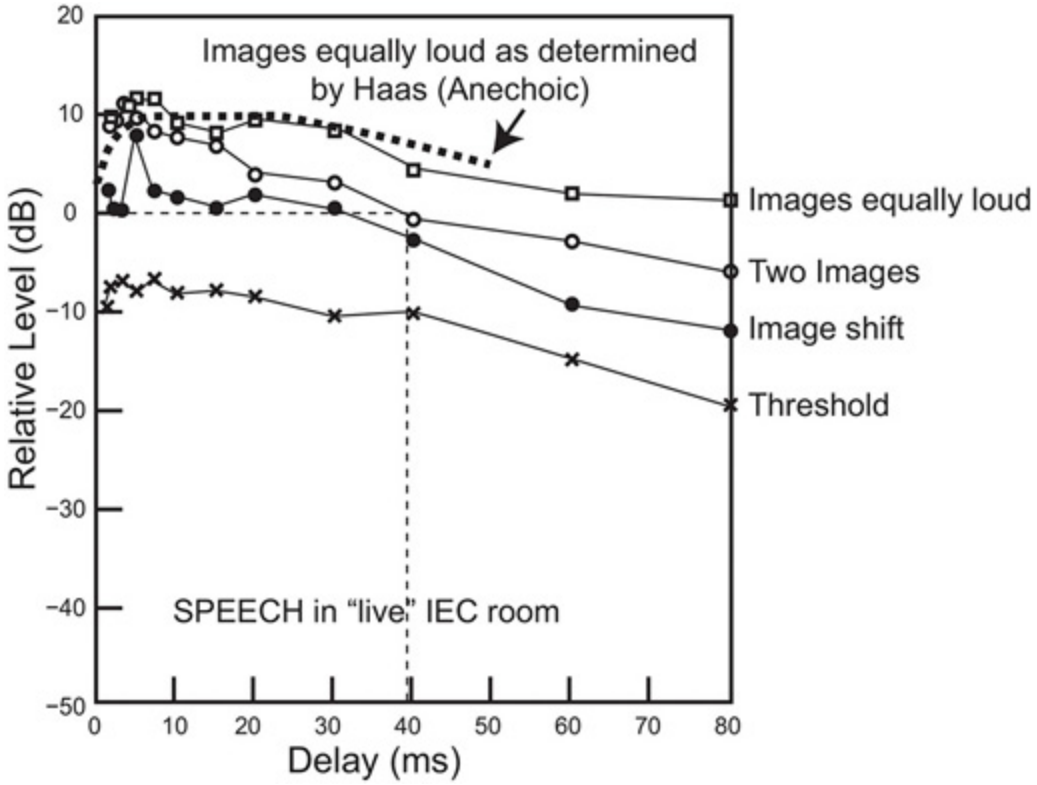


FIGURE 6.8 *The full set of thresholds, as shown in Figure 6.4, but here obtained while listening in a normally reflective room rather than an anechoic chamber. One listener (S.O.). Also shown is the precedence effect fusion interval for equal level direct and delayed sounds, the delay at which two images are just perceived. Unpublished data acquired during the experiments of Olive and Toole (1989). The Haas equal-loudness curve from Figure 6.2 is shown for comparison.* ⏎

The threshold curves become more horizontal; the fusion zone is extended in time when the sound is prolonged by reflected energy (repetitions) in the listening space. It was found that reverberation in the program had the same effect.

6.2.1 The "Family" of Curves in a Normal Room

Figure 6.8 shows a complete set of thresholds, like those shown in Figure 6.4 determined in an anechoic chamber, but here determined in the "live" IEC listening room. The curves are slightly irregular because the data were based on a small number of repetitions and only one, albeit very experienced, listener, Sean Olive. As expected, the curves all have a more horizontal appearance than for speech auditioned in an anechoic environment.

It is significant that all of the curves have the same basic shape from detection at the bottom, to the Haas-inspired equal-loudness curve at the top. The agreement of the "equally loud" image curves, one determined by Hass on a hemi-anechoic rooftop in 1949 (Figure 6.2) is good. Further confirmation is in the precedence effect fusion interval for equal level direct and delayed sounds, shown here to be 40 ms, compared to 30 ms determined by Haas. Both observations add confirmation to the observation in Figure 6.7 that, at short delays, small room reverberation is not an important factor in these determinations.

6.3 EFFECTS ON REAL AND PHANTOM IMAGES

A phantom image is a perceptual illusion resulting from summing localization when the same sound is radiated by two loudspeakers. It is natural to think that these directional illusions may be more fragile than those created by a single loudspeaker at the same location. This was tested.

The evidence shown in Figure 6.9 applies to the simple case of a single lateral reflection, simulated in a normally reflective room with a loudspeaker positioned along a side wall. When detection threshold and image-shift threshold determinations were done first with real and then with

phantom center images, in the presence of an asymmetrical single lateral reflection, the differences were insignificantly small. It appears that concerns about the fragility of a phantom center image are misplaced.

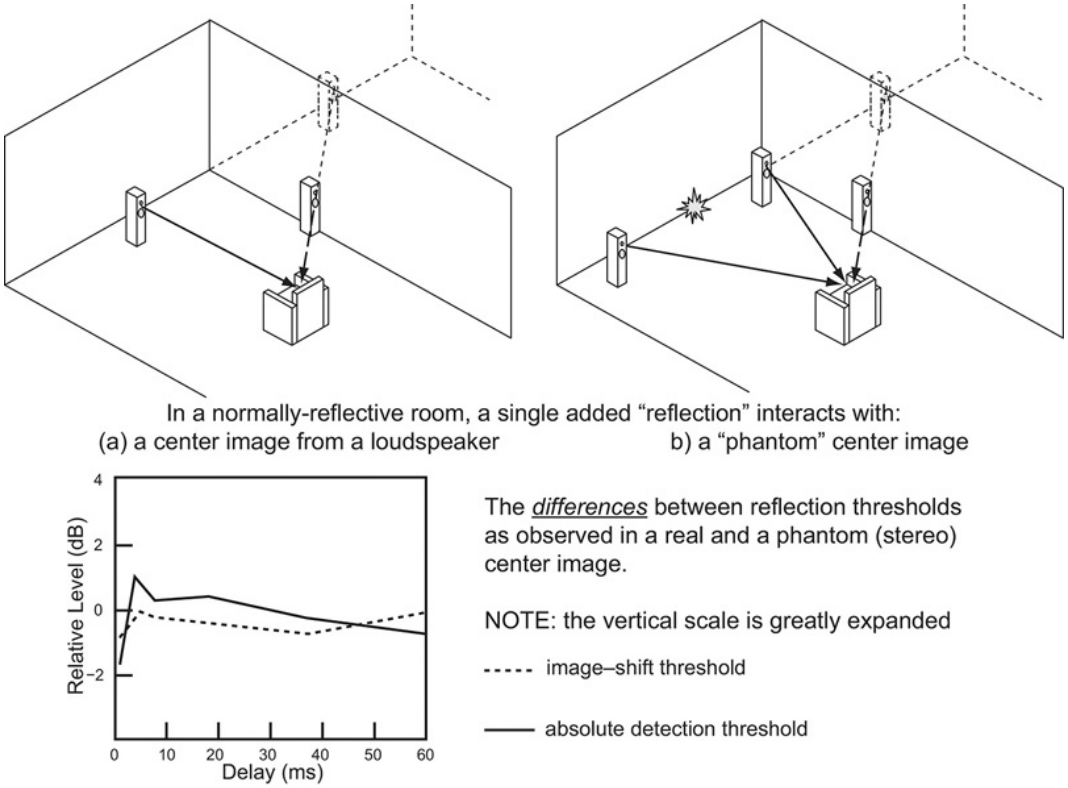


FIGURE 6.9 *An examination of how a real and a phantom center image respond to a single lateral reflection simulated by a loudspeaker located at the right side wall. The room was the "live" version of the IEC listening room used in other experiments. Note that the vertical scale has been greatly expanded to emphasize the lack of any consequential effect. The signal was speech.* ⏎

Examining the phantom image in transition from front to surround loudspeakers (±30° to ±110°), Corey and Woszczyk (2002) concluded that adding simulated reflections of each of the individual loudspeakers did not significantly change image position or blur, but it did slightly reduce the confidence that listeners expressed in the judgment.

6.4 RESULTS WITH MUSIC AND OTHER SOUNDS

It is useful to compare the shapes of the threshold contours determined by Barron (1971) who used an anechoic-chamber recording of Mozart's *Jupiter* symphony, with those shown earlier for speech, both in anechoic listening conditions. Figure 6.10 shows such a comparison, and it is seen that curves obtained using the anechoically recorded Mozart excerpt are much flatter than those for speech, even without concert hall reverberation.

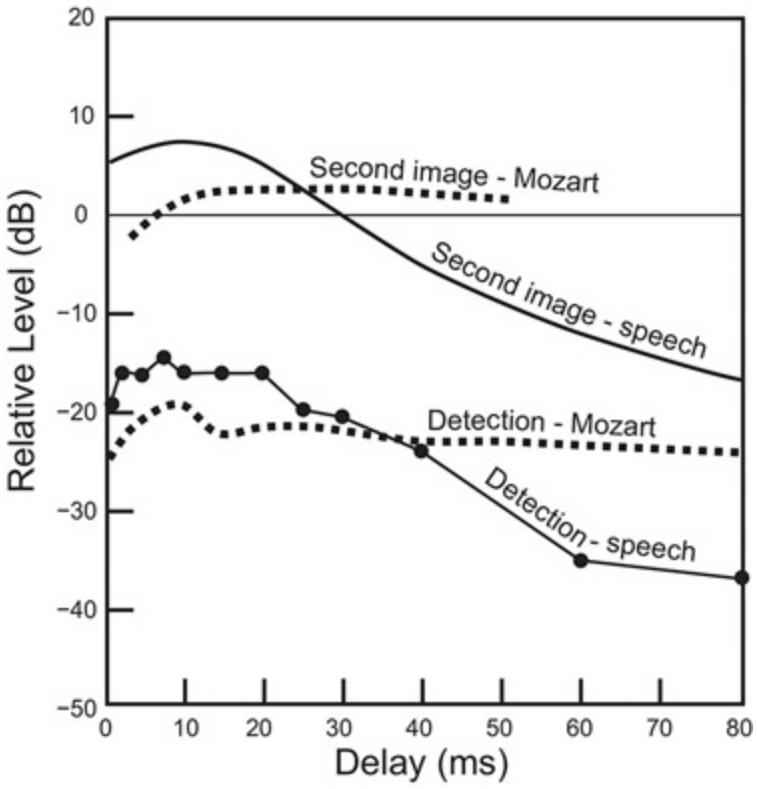


FIGURE 6.10 *Data from Figure 6.5(a) showing thresholds obtained using speech, and data from Barron, 1971, showing thresholds obtained using (anechoic) Mozart. The upper curve for music was described as that at which the "apparent source moved from direct sound loudspeaker towards reflection loudspeaker". This could be interpreted as being equivalent to the Olive and Toole "image shift" threshold, but the pattern of the data in the comparison suggests that it is more likely equivalent to the "second image" criterion. This is how it is labeled.* ⏎

Figure 6.10 shows two things. First, it appears that the slower paced, longer, notes in the music cause the threshold curves to be flatter than they are for the more compact syllables and consonant sounds in speech. This "prolongation" appears to be similar in perceptual effect to that occurring due to reflections in the listening environment (Figure 6.7). Second, it appears that the slope of the absolute threshold curve is similar to that of the second-image curve, something that was foreshadowed in Figure 6.4.

Figure 6.11 shows detection thresholds for sounds chosen to exemplify different degrees of "continuity", starting with continuous pink noise, and moving through Mozart, speech, castanets with reverberation, and "anechoic" clicks (brief electronically generated pulses sent to the loudspeakers). The result is that increasing "continuity" produces the kind of progressive flattening seen in Figures 6.7 and 6.10. The perceptual effect is similar if the "continuity" or "prolongation" is due to variations in the structure of the signal itself, or due to reflective repetitions added in the listening environment. In any event, pink noise generated an almost horizontal flat line, Mozart was only slightly different over the 80 ms delay range examined, speech produced a moderate tilt, castanets (clicks) with some recorded reverberation were even more tilted, and isolated clicks generated a very compact, steeply tilted threshold curve. Clicks showed detection of reflections at −50 dB at 30 ms delay, compared to −30 dB for castanets, and about −20 dB for speech, Mozart and pink noise.

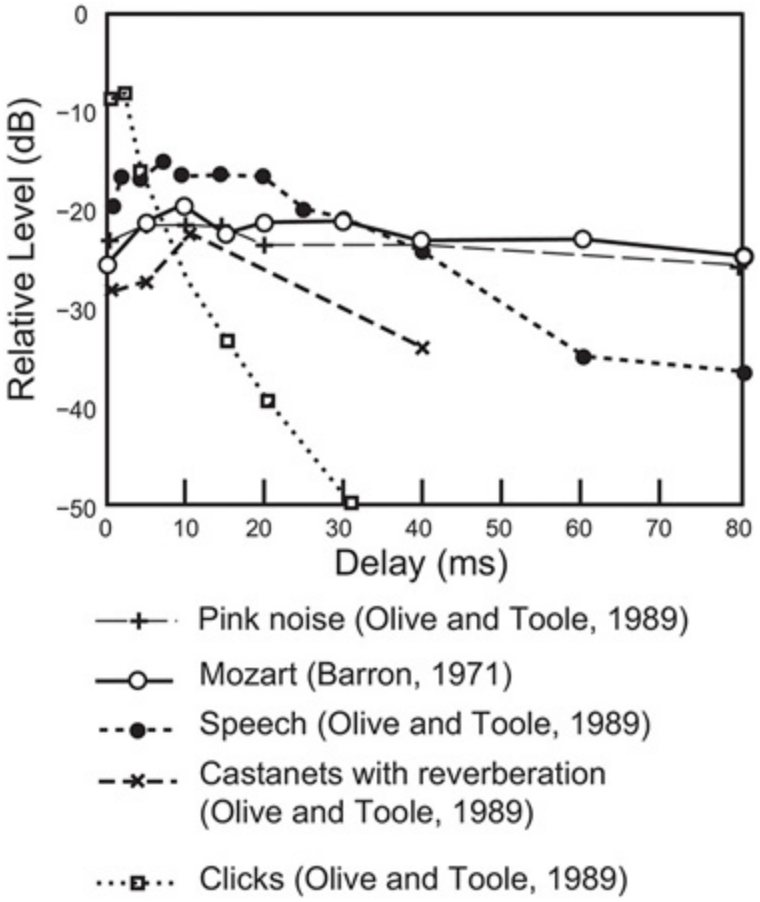


FIGURE 6.11 *Detection thresholds for a single lateral reflection, determined in an anechoic chamber for several sounds that exhibit different degrees of "continuity" or temporal extension.*

Assuming that the patterns seen in previous data for speech and Mozart apply to other sounds as well, Figure 6.12 shows a compilation and extension of data portraying detection thresholds and second-image thresholds for Mozart, speech and clicks. To achieve this, the second-image curve for clicks had to be "created" by elevating the click threshold curve by an amount similar to the separation of the speech and music curves. Absolute proof of this must await more experiments, but it is interesting to go out on a (strong) limb and speculate.

Looking at the 0 dB relative level line—where the direct and reflected sounds are identical in level—it can be seen that the precedence effect

interval for clicks appears to be just under 10 ms. According to Litovsky et al. (1999) this is consistent with other determinations (>10 ms), and the approximately 30 ms for speech is also in the right range (>50 ms). They offer no fusion interval data for Mozart, but it is reasonable to speculate based on the Barron data that it might be substantially longer then 50 ms.

Based on this evidence and that in Figure 6.5, if image quality is a concern, it would seem that acoustical treatment of naturally occurring small-room reflections is a debatable issue. The short fusion interval for clicks suggests that sounds like close-miked, unreverberated, percussion instruments might generate a second image, but only if there are no room reflections after the initial ones to extend the fusion zone.

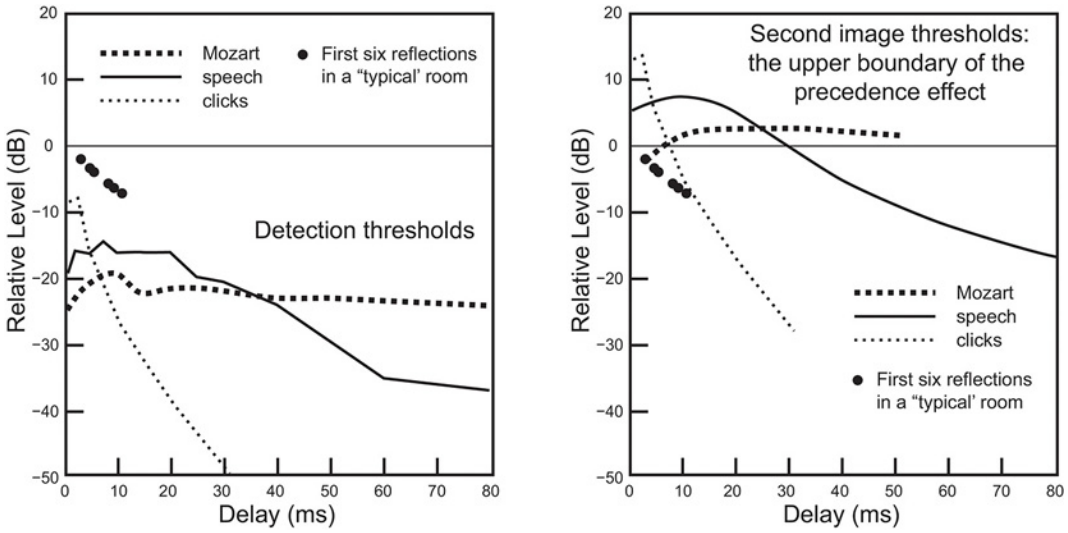


FIGURE 6.12 *Using data from Figure 6.10 and 6.11 this is a comparative estimate of the detection thresholds and the second-image thresholds (i.e. the boundary of the precedence effect) for clicks, speech and Mozart. The "typical room reflections" suggest that in the absence of any other reflections, the clicks are the only one of these sounds that approaches the point of being detected as a second image. However, normal room reflections would be expected to prevent this from happening because the threshold curve would be flattened (see Figure 6.7).* ⏎

Obviously, more research is necessary on this topic because in sound reproduction listeners wish to hear the reflections in the recording, not those of the small listening room. This is especially true in recording control rooms, which is why reflection-free zone (RFZ) rooms are created by absorbing or deflecting early listening room reflections. Evidence is scarce, but the practice began in the era of stereo, which is still with us of course, but now recordings are created in multichannel and immersive formats which are much more persuasive in delivering large space envelopment and spaciousness. With multiple sources of direct sound around the listener, it is probable that the small room acoustical contributions may be less significant. Section 5.8 discusses how recorded sounds of large spaces in multiple channels degrades our ability to hear flaws in loudspeakers. Do these sounds also mask aspects of the listening space? From the entertainment perspective, Hughes et al. (2016) have found that it is the larger space that is perceptually dominant.

Whether the small room reflections interfere with the recording process or not, the reality is that customers do not listen in RFZ rooms, so if translation is to occur, the master recordings must at some point be evaluated in typically reflective rooms.

6.5 REFLECTING, ABSORBING AND DIFFUSING SURFACES

Up to this point we have looked at some audible effects of single reflections when they appear in anechoic isolation and when they appear in the presence of room reflections. Obviously, hard flat surfaces are near perfect reflectors of sound, and in many situations, this is acceptable, even desirable. But some circumstances require controlling reflections, and there are two options for attenuating the energy in sound reflected from a surface: absorb the incident sound, or scatter it in many directions, sending only a fraction in the direction of the listener.

Irregular surfaces and objects, elegant ornamentation in older concert halls, have been important components in adding diffusion to performance

spaces where the precious sounds of live voices and instruments need to be blended and delivered to listeners distributed over large areas of large rooms. These are acoustically “live” spaces with long reverberation times—1 to 2 s being common. The hall is part of the performance.

Sound reproduction is a very different process. Typical listening rooms are small and much less reverberant, typically 0.3 to 0.5s. Reproducing systems have no shortage of sound power. It is possible to design the room interior using different surface treatments to meet a variety of objectives appropriate to stereo and multichannel systems. A highly diffuse sound field is not a requirement—nor is it possible in domestic spaces—but diffusers and scattering surfaces are useful devices to attenuate specific early reflections and to redirect the remaining sound energy in perceptually rewarding or inoffensive directions.

6.5.1 Absorbing High Frequencies Is Not Enough

Above the transition frequency fibrous (resistive) absorbers are widely used to reduce the energy in reflected sounds. Their performance is dominated by their thickness, becoming rapidly less effective when the thickness is less than about 1/4 wavelength because particle velocity diminishes to zero at the reflecting surface. For now, it is sufficient to know that resistive absorbers are, in effect, low-pass filters for reflected sounds; they attenuate the higher frequencies. Impulsive measurement methods tend to respond visually to the elimination of high frequencies, and this is often regarded as a measure of success.

Figure 6.13 shows some detection threshold measurements for sounds that have been low-pass filtered at various frequencies, from Olive and Toole (1989).

These limited data indicate that the effect of low-pass filtering has varied effects depending on the nature of the signal. For pulses at 1/s, speech and pink noise there was relatively little change in thresholds until the low-pass frequency reached 500 Hz.

Most of the spectrum was gone, yet the delayed sounds were detected at not very different levels, although the sound quality of what was being

detected was very different. For pulses at increased repetition rates, 5/s and 20/s, the thresholds progressively elevated as high-frequency energy was removed. The reduction in disturbance by later sounds was also noted by others, including Haas. This is an area that warrants further research, in part because it needs to be better correlated with physical measurements of events in the time domain, which will be discussed in Section 6.6.

6.5.2 Reflections From Smooth vs. Irregular Surfaces

If control of a specific reflection is required over a wide bandwidth, the necessarily thick resistive absorber might remove too much energy. Curved surfaces and irregular masonry surfaces have been used to scatter incident sounds in a variety of directions, thereby attenuating, not eliminating portions of the spectrum. Engineered surface diffusers are another option. With these devices incident sound energy is scattered over wide angles, attenuating the energy in any one direction, and adding multiple versions of it from each of the prominent surface features, spreading the energy out in time. Instinct says that both are useful behaviors, but one needs proof. Unfortunately, there seems to be very little. There are mathematical modeling exercises, many focused on performing spaces where increasing reflections/diffusion is the objective. In small listening rooms diffusion is not a goal. In small listening rooms, where they tend to be used as an alternative to absorbers, there is limited information and the subjective side of the scientific balance is substantially missing.

Cremer and Müller (1982) provide a limited but interesting perspective. Figure 6.14 shows a microphone picking up the direct sound from a loudspeaker and either a single large or three smaller reflections in rapid sequence. The middle layer of images displays sound pressure, showing the direct sound followed by the reflections. The bottom layer of images portrays what Cremer and Müller call an “ear-imitative” function, which is a simple attempt to show that the ear has a short memory that fades with time—a relaxation time. The point of this illustration is that events occurring within short intervals of each other can accumulate “effect”, whatever that may be. The sequence of three smaller reflections can be seen

to cause the “ear-imitative” function to progressively grow, although not to the same level as that for the single reflection.

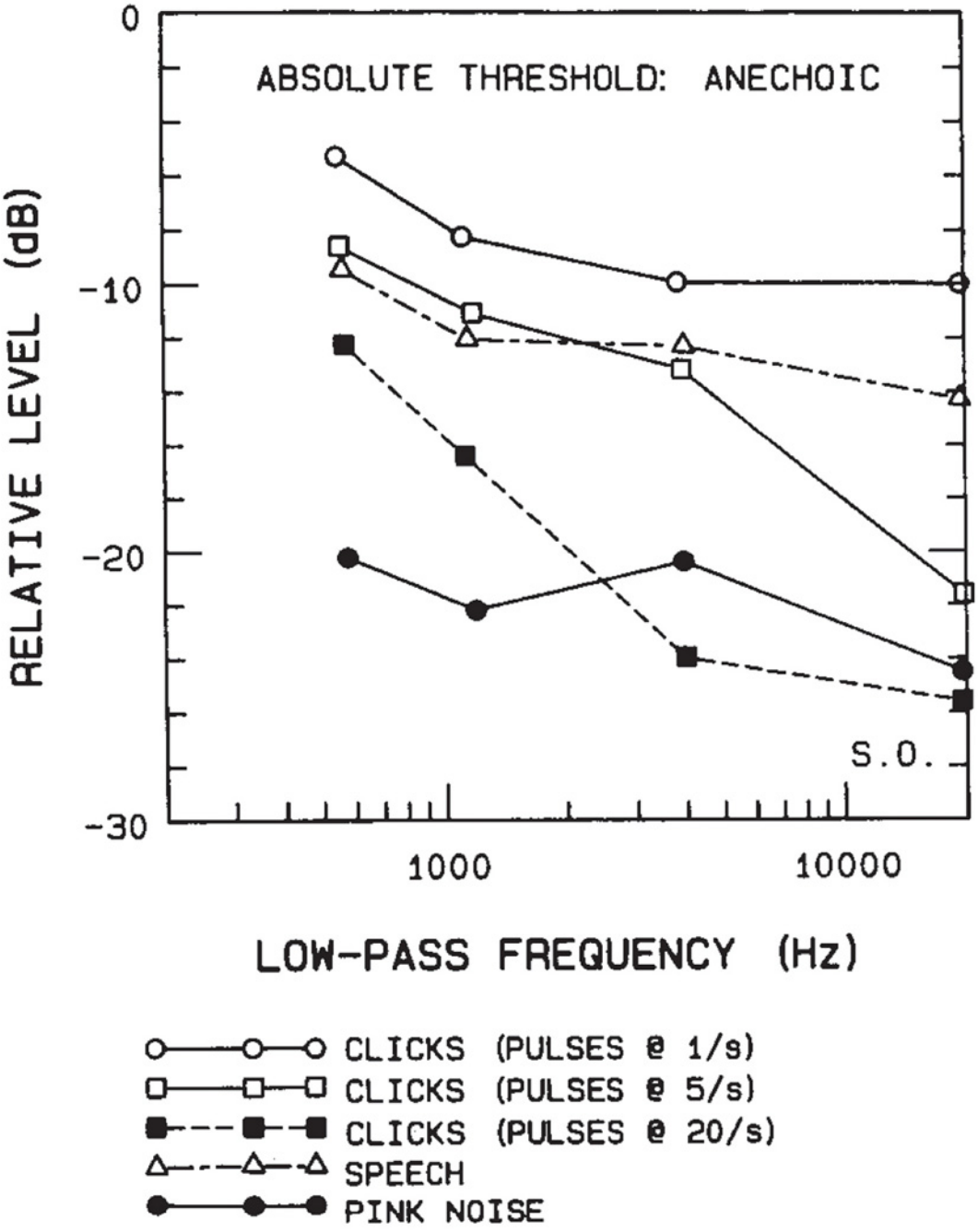


FIGURE 6.13 *Absolute thresholds for reflections arriving from 50° horizontal and 0° vertical at 4 ms delay. Signals range from discontinuous pulses at 1/s, to continuous pink noise. They have*

been presented broadband, and low-pass filtered at 4.4 kHz, 1.2 kHz and 500 Hz. Figure 16 from Olive and Toole (1989). ⏎

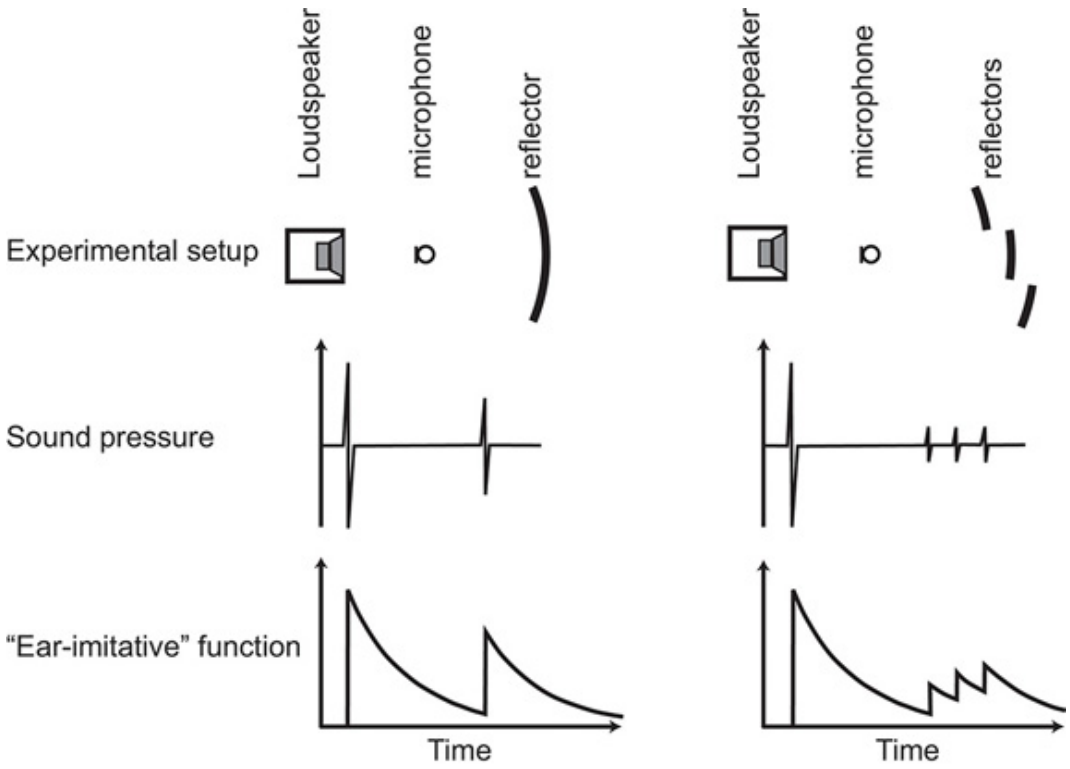


FIGURE 6.14 *A comparison of a single large reflection with a sequence of three lower-level reflections. From Cremer and Müller (1982, figure 1.16).* ⏎

However, when the authors conducted subjective tests in an anechoic chamber, they found that the sequence of three low-level reflections and the large single reflection were "almost equally loud". The message here is that if we believed the impulse response measurements, we might have concluded that by breaking up the large reflecting surface we had reduced the audible effects. This is one of the persistent problems of psychoacoustics. Human perception is usually non-linear, and technical measurements faithfully linear.

Angus (1997, 1999) compared large single, lateral reflections from a side wall with diffuse (i.e. multiple small) reflections from the same surface covered with scattering elements. There were no subjective tests, but

mathematical simulations showed some counter-intuitive results, namely that although the amplitudes of individual reflections were attenuated (as seen in an ETC), the variations in frequency responses measured at the listening position were not necessarily reduced. If the Cremer and Müller perceptual-summation effect is incorporated, the multiple smaller reflections seen in the ETC may end up being perceived as louder than anticipated. It is suggested, however, that a diffuse reflecting surface may make listening position less critical.

So, there are both subjective and objective perspectives indicating that breaking up reflective surfaces may not yield results that align with our instincts. It is another of those topics needing more investigation, especially controlled double-blind subjective evaluations—the difficult part. Meanwhile, there is an abundance of marketing claims and opinions.

6.6 INTERPRETING THE MEASURED EVIDENCE OF REFLECTIONS

6.6.1 Comb Filtering

It may be the name "comb filtering"—it sounds unpleasant. The name describes the visual appearance of the frequency response measured under some circumstances, a succession of deep notches—it looks disturbing to audio people anticipating flattish, smooth curves. A visit to the internet will find explanations, illustrations and some alarmist predictions of poor sound quality.

Comb filtering in sound fields is the result of acoustical interference between two sounds that arrive at the same point at different times. In normal rooms it is unavoidable because the direct sound from a source is followed by many reflected, delayed, versions of it. These combine, and depending on the wavelengths of the sounds, there may be summation (constructive interference) or cancellation (destructive interference). Consequently, the measured frequency response is very irregular. This is the "grass" seen in high-resolution room curves, the reason why room curves

tend to be spectrally smoothed—they look better. A concert hall can be viewed as an exceptionally complex comb filter, and yet we find it an especially pleasant place in which to listen. So, evidently, there is a difference between what is measured by a microphone and what is heard by two ears and a brain.

The classic explanation of comb filtering begins with the unrealistic situation of identical broadband direct and delayed sounds being combined. In the real world a delayed (reflected) sound will be lower in amplitude, having traveled farther and having encountered a reflecting surface of uncertain acoustical properties, so the spectrum may also be different. But for the purposes of the demonstration, when identical sounds combine, they sum perfectly at frequencies of constructive interference, and the amplitude doubles (6 dB). At frequencies of perfect destructive interference, the amplitude drops to zero. If the sounds contain all frequencies at constant amplitude a frequency response of the combined sounds will show a series of peaks and dips related to the frequency/wavelength of the sounds and the delay. On a linear frequency scale, the pattern is regular across the frequency range, hence the name "comb". However, in audio we plot on a logarithmic frequency scale, so the peaks and dips progressively crowd together as frequency increases. Figure 6.15 illustrates this situation.

Figure 6.15(a) shows a logarithmic frequency response measurement for equal level direct and delayed sounds with a delay of 1 ms. The first cancellation dip occurs at the frequency at which 1 ms is one-half wavelength, which is 500 Hz, where the direct and delayed sounds add perfectly. The second is at the odd-integer frequency of 1500 Hz, etc. The peaks occur at even-order integer multiples of 500 Hz.

The solid black line describes the situation with equal-level direct and delayed sounds, with which perfect summation (+6 dB) and cancellation ($-\infty$) occur. The delay is very small, only 1 ms, which yields alarmingly large fluctuations in the middle frequencies. This is the illustration found (with variations) in a number of popular texts describing the phenomenon of comb filtering, and it creates an impression that this is a major problem in audio. If this is a scare tactic, it works. This certainly looks ugly.

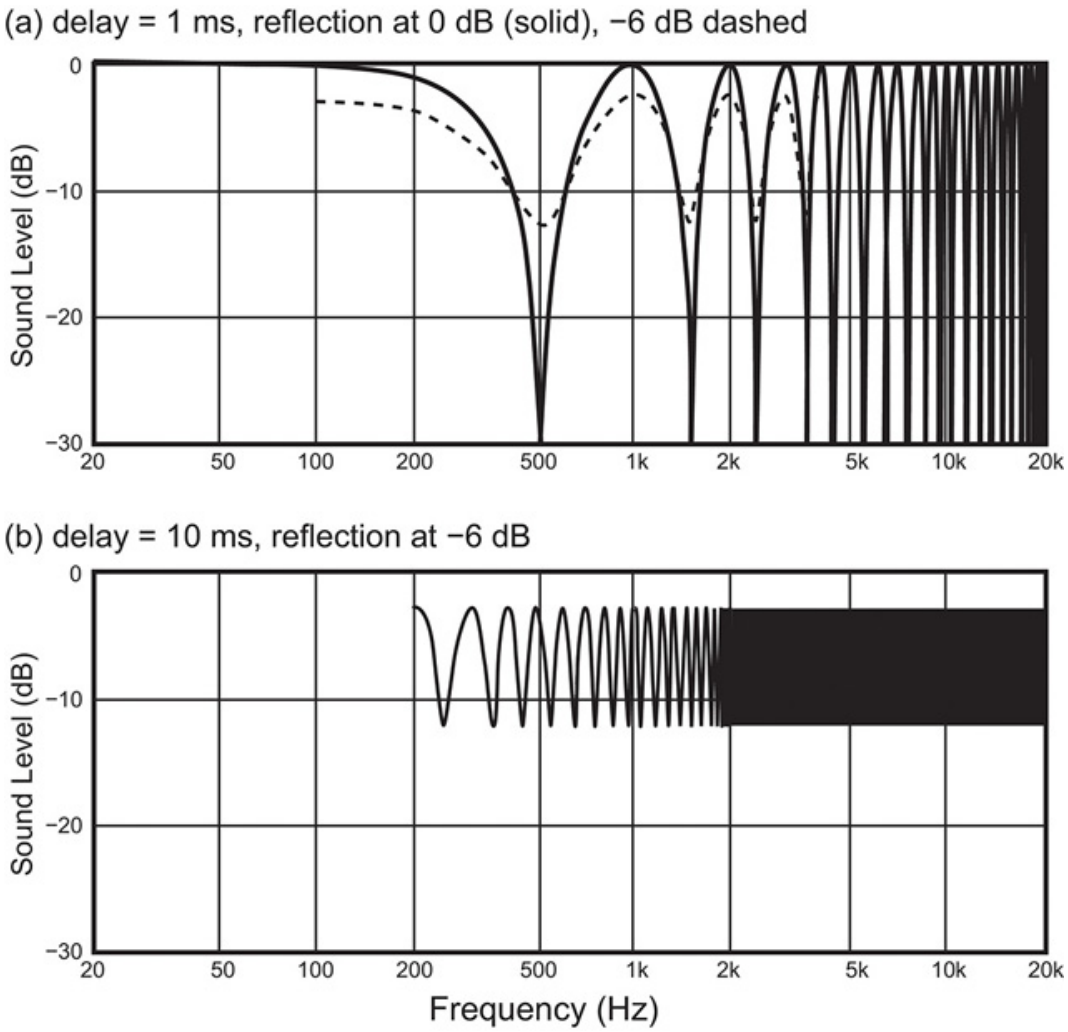


FIGURE 6.15 *Illustrations of idealized comb filtering in frequency response curves shown on a logarithmic frequency scale. In (a) the solid line shows the interference pattern when the delay is 1 ms and the direct and delayed sounds have identical amplitudes. The dashed line shows what happens when the delayed sound is attenuated by 6 dB. In (b) the delay is increased to 10 ms, moving the entire interference pattern down the frequency scale and the reflection is a realistic −6 dB. Below 200 Hz the curve is terminated because in a small room, performance in this frequency range will be dominated by room resonances/standing waves.*

However, in listening rooms the delayed sounds are reflections, and these will occur at much longer delays and they will be attenuated in amplitude by propagation loss (about −6 dB/double-distance) and absorption at the reflecting surfaces (can be anything from near zero for a hard flat surface to several dB for a deep, fluffy resistive absorber). The dashed line in Figure 6.15(a) shows the result of a reflection reduced in amplitude by 6 dB. The peaks are a little lower and the depth of the dips is greatly reduced. It looks much less alarming. This is important.

At a delay more typical of events in listening rooms, the whole pattern moves down the frequency scale. Figure 6.15(b) shows a reflection arriving after a 10 ms delay (corresponding to a path length that is 11.3 ft (3.4 m) longer than the direct-sound path. The corresponding sound level might be 6 dB lower than the direct sound. Obviously, the large undulations have moved to the bass region—below the transition frequency in small listening rooms, where sound propagation is dominated by wave-acoustic phenomena. We can ignore these effects as they will be merged with, and indeed swamped by, the standing-wave/room resonances active in this frequency region. At frequencies above about 200 Hz, it can be seen that the comb filtering undulations are so closely packed together that they cease to be the audible problem suggested by Figure 6.15(a). For longer delays, the density of the undulations is even greater.

Only very high-resolution measurements would reveal these details, normal 1/3- or 1/6-octave band curve smoothing would reveal none of this detail. What would typically be seen would be a single line that might suggest the first undulations at the low end of the frequency scale, gradually showing an inability to reveal the full depth of the dips, and from some frequency upwards simply reverting to a straight horizontal line following the top of the black area.

All of this assumes a “steady state” sound. Speech and music have many transient and short-duration decaying components, that would interact differently because they have no “steady-state” characteristics—the direct sound may be in decline before the reflected version arrives. These curves do *not* depict how a listener would perceive a normal sound with a complex

frequency and time-domain structure. Steady-state measurements do not show what we hear in rooms. But reflections exist, and their presence can be confirmed in other kinds of measurements, like energy-time curves (ETC). Are they an audible problem?

At the beginning of this chapter it was noted that when listening to a single reflection in an anechoic chamber the dominant subjective impression was one of spaciousness. If one measured what was being heard it would show evidence of acoustical interference comb filtering. Viewed from this perspective what we hear in any room is a complex set of comb filters and we find pleasure in listening in these spaces. How can something that looks so bad sound so good? The explanation is in Figure 1.5—we listen with two ears and a brain, the binaural hearing system, and we measure with an omnidirectional microphone. Delayed sounds arriving from directions other than the direct sound are perceived as components of spaciousness. But even those sounds that arrive from the same direction as the sound source are perceptually discriminated against—see Figure 6.6.

This is not to say that all reflected sounds are beneficial. Most are innocent information about the room we are in, and it is possible that a few might be excessively strong. But, it is improbable that the objectionable perceptions are caused by comb filtering.

The worst situation for the audibility of comb filtering is when the summation occurs in the electrical signal path, as occasionally happens in live events, and in broadcasts, when a signal accidentally gets routed so that it combines with a delayed version of itself. We hear these occasionally in news broadcasts—it sounds as if the outputs of two microphones (among the sometimes massive collections) separated by small distance were blended. The audible coloration is not subtle. The problem is easily heard because in the listening room the direct sound and all reflected versions of it contain the same interference pattern. Clark (1983) describes an interesting experiment that demonstrates it in measurements (see figure 7.3 in the third edition, or OR Chapter 6).

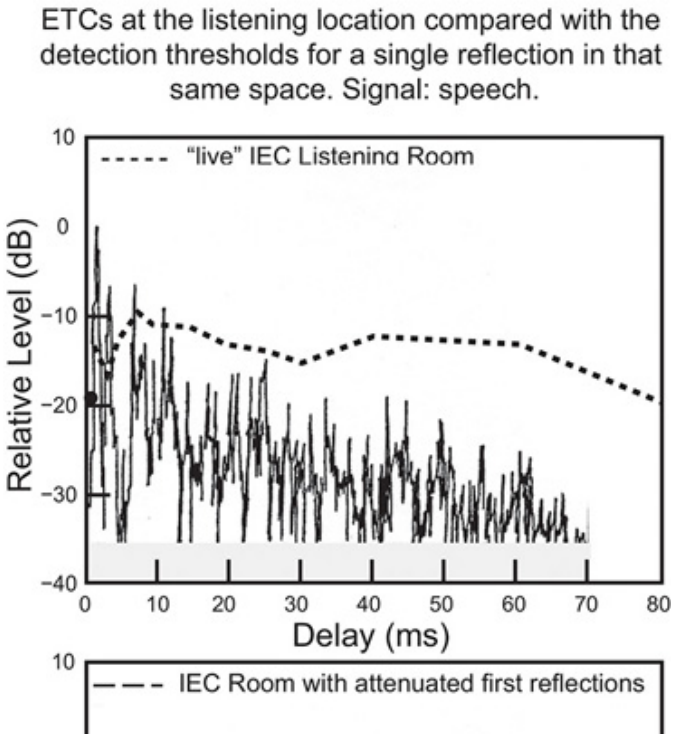
Large public-address/sound reinforcement loudspeaker arrays, with many units spread out over a large area, all radiating the same sounds, obviously

have the potential to be problematic. Dealing with these issues is fundamental to the design of such arrays, and elaborate measurements of the system components are used with mathematical models to predict the three-dimensional sound field radiated by arrays, to anticipate and lessen these effects.

Another difficult situation is one in which there is only a single dominant reflection arriving from close to the same direction as the direct sound. In a control-room context, this could be a console reflection in an otherwise acoustically damped room. The very short delay in this situation is similar to that shown in Figure 6.15(a).

6.6.2 Energy-Time Curves

Figure 6.16 shows a direct comparison of the absolute detection thresholds with the ETC (energy-time curve) measured in each of the three test environments used in the experiments of Figure 6.7. There are large variations in level of the reflections, but relatively small changes in the detection thresholds within the first 30 ms or so. What explains the apparent disagreement?



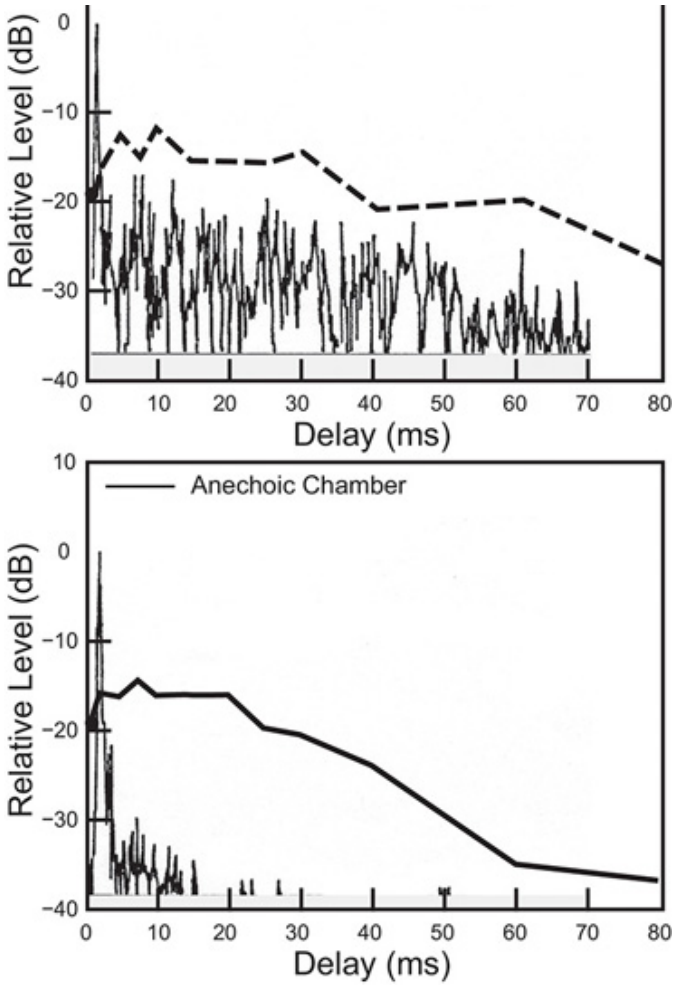


FIGURE 6.16 *A comparison of the absolute thresholds shown in Figure 6.7 with measured energy-time curves (ETCs) for the three spaces within which the tests were done. All data from Olive and Toole (1989).* ⏎

It seems obvious to look at reflections in the time domain, in a "reflectogram" or impulse response, a simple oscilloscope-like display of events as a function of time or, the popular alternative, the ETC (energy-time curve). In such displays, the strength of the reflection would be represented by the height of the spike. However, the height of a spike is affected by the frequency content of the reflection, and, without pre- or post-processing, time-domain displays are "blind" to spectrum. The measurement displays no information about the frequency content of the

sound it represents. Only if the spectra of the sounds represented by two spikes are identical can they legitimately be compared.

Let us take an example. In a very common room acoustic situation, suppose a time-domain measurement reveals a reflection that it is believed needs attenuation. Following a common procedure, an absorber is placed at the reflection point. In this example it is a theoretically perfect absorber—the kind that does not exist—in that it removes all sound above a certain frequency, in this case 500 Hz. A new measurement is made and—behold!—the spike has gone down. Success, right? Maybe not.

In a controlled situation, Olive and Toole (1989) performed a test intended to show how different measurements portrayed reflections that, subjectively, were adjusted to be at the threshold of detection. So, from the listener's perspective, the two reflections that are about to be discussed are the same—just at the point of audibility or inaudibility.

The results are shown in Figure 6.17. At the top, the (a) graphs are waterfall diagrams displaying events in three dimensions, amplitude vertically, time moving forward, and frequency horizontally. At the rear is the direct sound, the next event in time is an intermediate reflection and at the front is the second reflection, the one that we are interested in. It can be seen that the second reflection is broadband in the left-hand diagram and that frequencies above 500 Hz have been eliminated in the right-hand version. When that particular "layer" of the waterfall is isolated, as in the (b) displays, the differences in frequency content are obvious. The amplitudes are rather similar, although the low-pass filtered version is a little higher, which seems to make sense considering that slightly over 5 octaves of the audible spectrum have been removed from the signal. Recall that these signals have been adjusted to produce the same subjective effect—a threshold detection—and it would be logical for a reduced-bandwidth signal to be higher in level.

In contrast, the (c) displays, showing the ETC measurements were telling us that there might be a difference of about 20 dB in the opposite direction; the narrow-band sound is shown to be lower in level. Obviously, this

particular form of the measurement is not a good correlate with the audible effect in this test.

The message is that we need to know the spectrum level of reflections in order to be able to gauge their relative audible effects. This can be done using time-domain representations, like ETC or impulse responses but it must be done using a method that equates the spectra in all of the spikes in the display, e.g. bandpass filtering. Examining the “slices” of a waterfall would also be to the point, as would performing FFTs on individual reflections isolated by time windowing of an impulse response. Such processes need to be done with care because of the trade-off between time and frequency resolution. It is quite possible to generate meaningless data.

All of this is especially relevant in room acoustics because acoustical materials, absorbers and diffusers, routinely modify the spectra of reflected sounds. Whenever the direct and reflected sounds have different spectra simple broadband ETCs or impulse responses are not trustworthy indicators of audible effects.

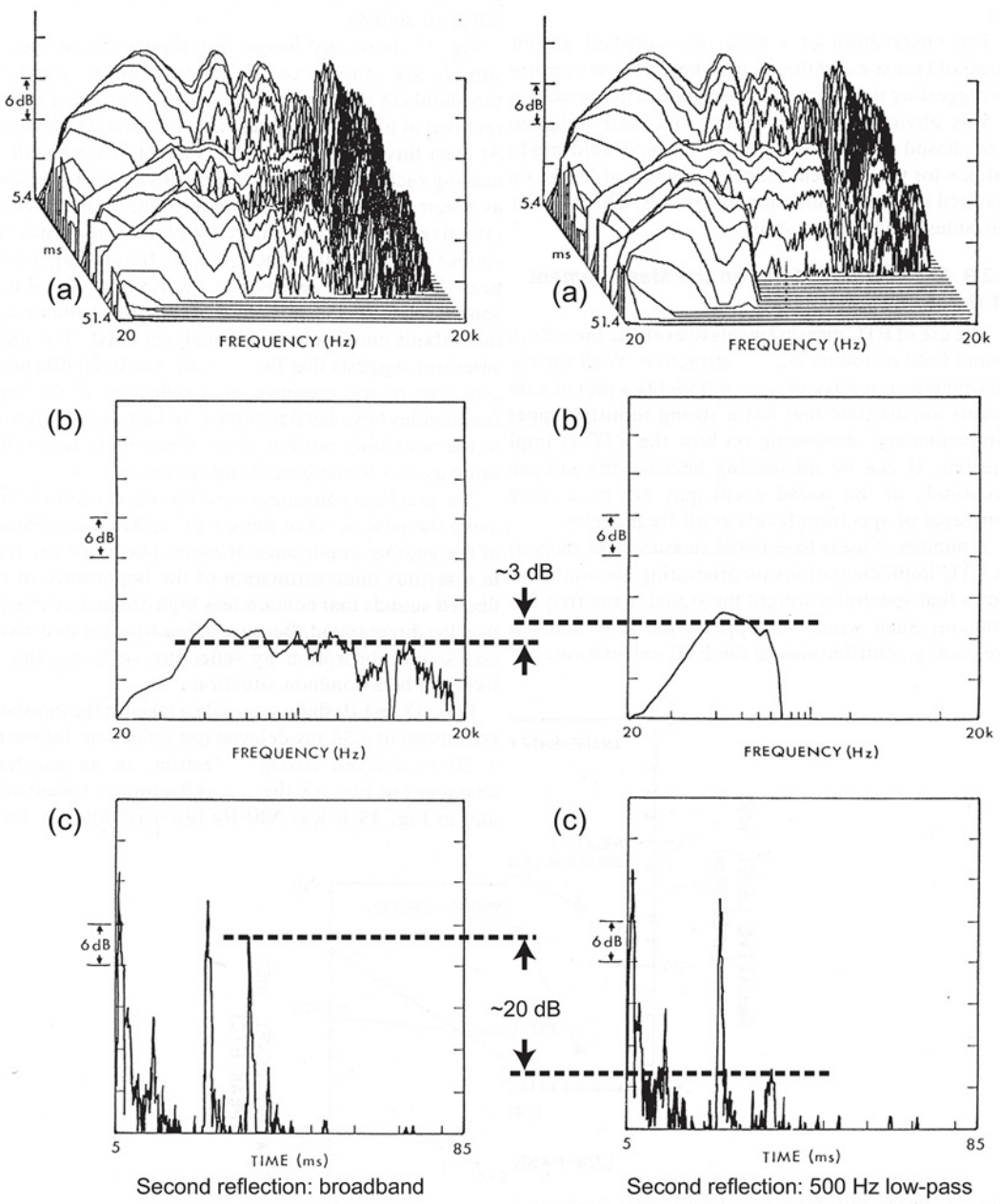


FIGURE 6.17 *The left column of data shows results when the second of a series of reflections was adjusted to the threshold of detection when it was broadband, the right column shows comparable data when the reflection was low-pass filtered at 500 Hz. (a) shows the waterfall diagram, (b) the spectrum of the second reflection taken from the waterfall, and (c) shows the ETC measured with a Techron 12 in its default condition (Hamming windowing). The signal was speech. The horizontal dotted lines*

are "eyeball" estimates of reflection levels. From Olive and Toole (1989) figures 18 and 19. ⏎

REFERENCES

Angus, J. A. S. (1997). "Controlling Early Reflections Using Diffusion", *102nd Convention, Audio Eng. Soc*., Preprint 4405. ⏎

Angus, J. A. S. (1999). "The Effects of Specular Versus Diffuse Reflections on the Frequency Response at the Listener", *106th Convention, Audio Eng. Soc*., Preprint 4938. ⏎

Barron, M. (1971). "The Subjective Effects of First Reflections in Concert Halls—the Need for Lateral Reflections", *J. Sound Vib*., vol. 15, pp. 475–494. ⏎

Benade, A. H. (1985). "From Instrument to Ear in a Room: Direct or via Recording", *J. Audio Eng. Soc*., vol. 33, pp. 218–233. ⏎

Blauert, J. (1996). *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, MA. ⏎

Brown, A., Stecker, C. and Tollin, D. (2015). "The Precedence Effect in Sound Localization", *J. Assoc. Res. Otolaryngol*, vol. 16, no. 1, February, pp. 1–28. www.ncbi.nlm.nih.gov/pmc/articles/PMC4310855/#CR78 ⏎

Burgtorf, W. (1961). "Untersuchungen zur wahrnehmbarkeit verzögerter schallsignale", *Acustica*, vol. 11, pp. 97–111. ⏎

Clark, D. L. (1983). "Measuring Audible Effects of Time Delays in Listening Tests", *74th Convention, Audio Eng. Soc*., Preprint 2012. ⏎

Corey, J. and Woszczyk, W. (2002). "Localization of Lateral Phantom Images in a 5-Channel System With and Without Simulated Early Reflections", *Audio Eng. Soc., 113th Convention*, Preprint 5673. ⏎

Cremer, L. and Müller, H. A. (translation by T. J. Schultz) (1982). *Principles and Applications of Room Acoustics*, Vol. 1 & 2, Applied Science Publishers, London. ⏎

Devantier, A. (2002). "Characterizing the Amplitude Response of Loudspeaker Systems", *113th Convention, Audio Eng. Soc*., Preprint

5638. ⏎
Gardner, M. (1968). “Historical Background of the Haas and/or Precedence Effect”, *J. Acoust. Soc. Am.*, vol. 43, pp. 1243–1248. ⏎
Gardner, M. (1969). “Image Fusion, Broadening, and Displacement in Sound Localization”, *J. Acoust. Soc. Am.*, vol. 46, pp. 339–349. ⏎
Gardner, M. (1973). “Some Single- and Multiple-Source Localization Effects”, *J. Audio Eng. Soc.*, vol. 21, pp. 430–437. ⏎
Haas, H. (1949, 1972). “The Influence of a Single Echo on the Audibility of Speech”, Doctoral dissertation, University of Göttingen. Reprinted in: *J. Audio Eng. Soc.*, vol. 20, pp. 146–159, 1972. A reprint of a 1949 translation of Haas’ PhD dissertation. ⏎
Hughes, R., Cox, T., Shirley, B. and Power, P. (2016). “The Room-in-room Effect and Its Influence on Perceived Room Size in Spatial Audio Reproduction”, *141st Convention Audio Eng. Soc.*, Preprint 9621. ⏎
Litovsky, R. Y., Colburn, H. S., Yost, W. A. and Guzman, S. J. (1999). “The Precedence Effect”, *J. Acoust. Soc. Am.*, vol. 106, pp. 1633–1654. ⏎
Lochner, J. P. A. and Burger, J. F. (1958). “The Subjective Masking of Short Time Delayed Echoes by Their Primary Sounds and Their Contribution to the Intelligibility of Speech”, *Acustica*, vol. 8, pp. 1–10. ⏎
Meyer, E. and Schodder, G. R. (1952). “On the Influence of Reflected Sound on Directional Localization and Loudness of Speech”, *Nachr. Akad. Wiss. Gottingen, Math. Phys. Klasse IIa*, vol. 6, pp. 31–42. ⏎
Olive, S. E. and Toole, F. E. (1989). “The Detection of Reflections in Typical Rooms”, *J. Audio Eng. Soc.*, vol. 37, pp. 539–553. ⏎
Seraphim, H.-P. (1961). “Über die wahrnehmbarkeit mehrerer rückwürfe von sprachschall”, *Acustica*, vol. 11, pp. 80–91. ⏎
Shinn-Cunningham, B. G. (2003). “Acoustics and Perception of Sound in Everyday Environments”, *Proc. 3rd Int. Workshop on Spatial Media*, Aisu-Wakamatsu, Japan. http://cns.bu.edu/~shinn/pages/RecentPapers.html ⏎
Toole, F. E. (2006). “Loudspeakers and Rooms for Sound Reproduction—A Scientific Review”, *J. Audio Eng. Soc.*, vol. 54, pp. 451−476. ⏎

Toole, F. E. (1990). “Subjective Evaluation: Identifying and Controlling the Variables”, *8th International Conference, Audio Eng. Soc*., Paper C-1. ⏎

CHAPTE
R 7

Space: A Motivator for More Channels

Floyd Toole

DOI: 10.4324/9781003477495-8

7.1 SPACE IS FUNDAMENTAL: REFLECTIONS ARE IMPORTANT

Monophonic reproduction delivers the basic music: melody, rhythm, vocal and instrumental content, and a hint of space in the form of reverberation. Reverberation in that case is a time-domain phenomenon because it emerges from a single loudspeaker—there is no sense of being in the same space as the performers, no envelopment.

Stereo adds a channel and provides a soundstage, with performers displayed between the loudspeakers. There is also a sense of space, totally depending on the goal of the recording artists and engineers. However, it is often described as listening through a hole in a wall into the concert hall,

even for the single listener in the prime location. The loudspeakers and listening room are often blamed for some of the disappointment.

Multichannel audio is a major advance, because it can deliver direct sounds and recorded reflections from different directions and at the long delays consistent with large venues. It moves the listener closer to what it might have been like at a live performance in a concert hall, and other performing spaces appropriate to the musical genre, like jazz or folk music. But most popular music is created in a recording control room, and many spatial illusions are possible, some realistic, some abstract. It is a powerful delivery system.

The science explaining what and why we hear directional and spatial qualities in large and small rooms is still developing. What follows is an assembly of scientific fragments that collectively may assist in explaining what we hear in our entertainment systems.

Griesinger (1989, 1997, 1998, 1999) has spent more time than most people analyzing spatial effects in concert halls and in their domestic counterpart, multichannel audio systems. His insights are thought-provoking; the 1997 paper seems to be a good summary. In the context of small listening rooms, Griesinger (2001) points out that, because most of the reflected sound energy occurs in the first 50 ms the most that this can achieve, perceptually, is what he calls ESI (early spatial impression). Such a spatial impression is predominantly frontal, "closely associated with the direct sound". Multiple reflections in this interval, and beyond, are not perceived as discrete events, but rather as a "fully enveloping surround" accompanying a sharply defined source, although musical sound having "slow note onsets—such as legato strings—… produces considerable source broadening (ASW). However, the spatial impression remains frontal" (Griesinger, 1997).

7.1.1 Perceptually Preferred Amounts of Reflected Sound

It is educational to review some of the pioneering research into the kinds of reflected sounds that are most preferred by listeners. What we learn may be surprising. For example: what happens when listeners are allowed to choose

the level of a single reflection, based on what they perceive as a sense of pleasantness—a preference?

Ando (1977) has provided some answers. Figure 7.1 shows his findings for speech, which are superimposed on data from Chapter 6 for perspective.

Figure 7.1 shows levels for a single delayed sound, from a horizontal angle of 36°, that listeners reported as enhancing the sound of speech. The early reflections in typical real rooms (the black dots) are all lower in level and earlier in time than the levels preferred by listeners. Multichannel audio is needed to provide added, stronger (by 7 to 10 dB) and later (by up to 30 ms), reflections for our listening pleasure (Ando, 1977, 1998). Note that the preferred levels just avoid the "second image" curve, indicating that the preferred reflections were all within the precedence effect fusion zone, thereby not generating distracting second images. The inevitable conclusion is that, in listening to live speech, and for a single loudspeaker reproducing speech, individual room reflections are not problems. In fact, they are not loud enough or late enough to put the speech into a pleasing acoustical setting. However, it is also evident that listeners do not wish to hear any secondary images—one person speaking is enough.

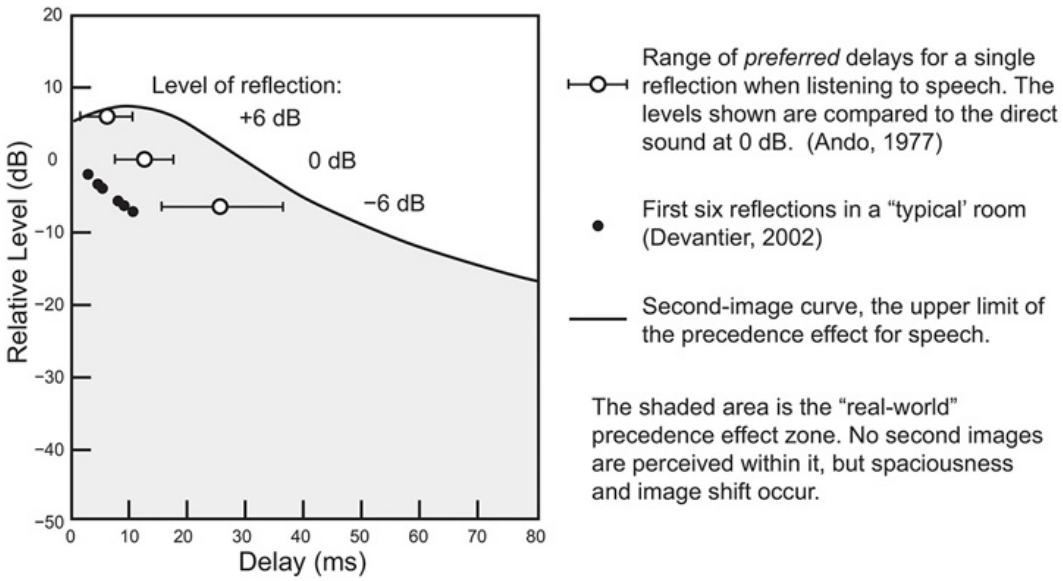


FIGURE 7.1 *With speech as the signal, the horizontal bars show delays at which listeners expressed a preference for the perceived sound experience when a single reflection was added*

from a horizontal angle of 36°. Results are shown for three sound levels relative to the direct sound. Adapted from Ando (1977, figure 7). The precedence effect and typical room reflection data are from Figure 6.5. ⏎

Unanswered in this study is the effect of the multiple reflections that occur in real spaces which, in combination, may be louder than the individual components. See Section 6.5.2 for discussion. The same comment applies to the following data.

Figure 7.2 shows experimental results from a similar experiment when the test sound was classical music. This figure needs to be examined in parts. First, three horizontal bars, at +6 dB, 0 dB and −6 dB show the delays at which listeners preferred single lateral reflections added to anechoic recordings of classical music at three different sound levels (Ando, 1977, 1998). Second, intermingled with these data are thin dashed lines indicating the levels at which different percentages of listeners preferred the addition of the single lateral reflection, again with classical music (Ando, 1985). High-level long-delayed reflections are not as desirable as those with shorter delays and lower levels. However, all preferred levels are much higher in amplitude and later in time than the natural reflections provided by small rooms, indicated by the six dots. Small room reflections are simply not strong enough or late enough to be useful.

In contrast to the results using speech, listeners in these tests did not avoid second images. The upper boundaries of the precedence effect fusion zones for speech and Mozart are shown and listeners demonstrated strong preferences for reflected sounds close to and outside the boundaries of the precedence effect. The subjectively preferred reflections were up to about 15 dB higher than natural room reflections, with delays up to 60 ms longer. Not to be forgotten is the fact that the sense of spaciousness has grown at all levels above threshold and continues to grow at levels outside the precedence zone. There are two perceptual phenomena involved: second images and increased spaciousness. It is not clear which, if either, is dominant in this instance. With classical music, hearing second images—

reflected versions of the direct sounds—appeared not to be offensive. The large reflecting surfaces in concert halls may have set a perceptual precedent. Spaciousness is a positive attribute, and multichannel audio is the enabler.

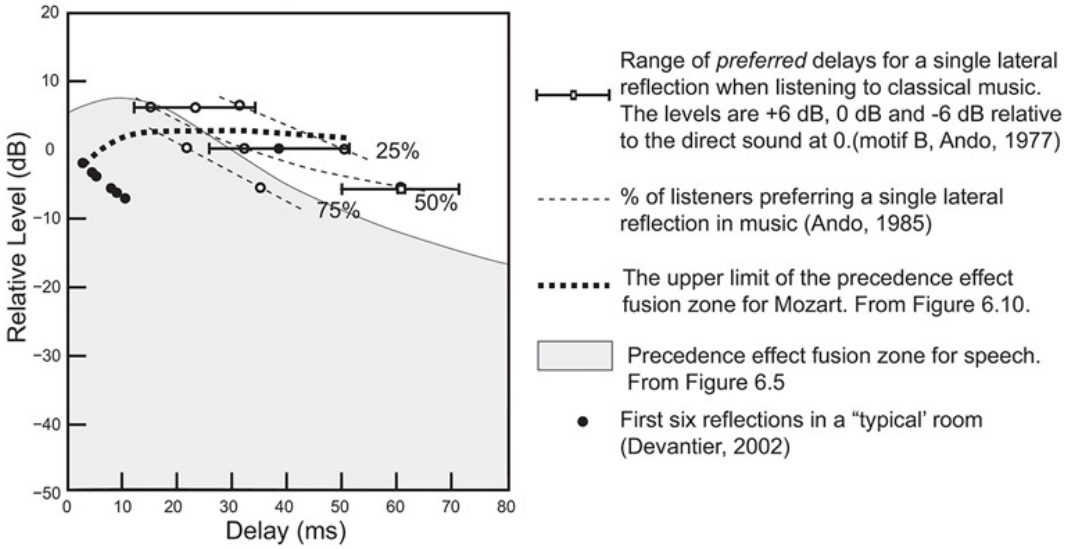


FIGURE 7.2 *A comparison of preferred levels for a single lateral reflection when listening to classical music. The estimated range of the precedence effect is shown.*

To be noted is that a measurement of what was heard in these two experiments would have revealed unambiguous comb filtering, a direct sound added to a delayed sound. Figure 1.5 explains why listeners liked it.

In his pioneering investigations of reverberation over 100 years ago, Sabine experimented with differing amounts of absorption in rooms of different sizes, asking listeners for opinions about how a live piano sounded. In rooms of typical living-room, home-theater sizes (2600 to 7400 ft^3/74 to 210 m^3) the most favored conditions were rooms having reverberation times in the range 1 to 1.1 s (reported in Cremer and Müller, 1982, p. 528). These had to be sparsely furnished rooms to achieve such high RTs, meaning that reproducing a close-miked recording of a piano in a normally furnished domestic room (RT 0.3 to 0.5 s) would be less than fully rewarding without added reverberation in the recording.

All these experiments convey the message that listeners prefer reflections at sound levels and delays than cannot be provided by small listening rooms. We need to have louder and later "reflected" sounds in the recordings, reproduced from multiple loudspeakers in the optimum locations. The missing information is: from which direction should these reflected sounds come?

These are observations based on simple experiments. Real-world circumstances are much more complex. And important questions emerge:

> If listeners prefer to hear reflections that are louder and later than those generated by the listening room, to what extent does the listening room detract from recordings that include the desired reflections? Does the number of channels delivering the desired reflections alter the perceptions of the listening room reflections?

In the extreme, would a many-channel immersive system reduce the demands on the acoustical properties of listening spaces? Hughes et al. (2016) provides some data, indicating that the impression of a recorded space dominates when it is larger than the playback space. This is good for home listening. The reverse is also true, confirming the difficulty of reproducing the perception of a small space while listening in a large one. Cinemas have trouble conveying intimate conversations that should have no room sound.

7.1.2 Perceptually Preferred Directions for Reflections

Concert hall acoustics provide an important clue, as the evidence favors lateral reflections from side walls. But independent investigations have provided supporting data. In the following tests interaural cross-correlation (IACC) is determined for sounds arriving at a listening location from different horizontal directions relative to 0°, directly forward. IACC is a statistical measure that correlates with perceptions of spaciousness and

envelopment. The lower the IACC, the more different the sounds at the two ears are, the greater the spatial impression.

Bringing the discussion closer to the small room topic, where reflections within the room are most capable of influencing ASW/image broadening, there is the relevant data from Hidaka et al. (1997), who concluded that the apparent source width (ASW) was most strongly influenced by reflected sounds in the octave bands 500, 1000 and 2000 Hz (embracing frequencies from about 350 to 2800 Hz) occurring within the first 80 ms. They created the measure $IACC_{E3}$ to reflect this. In Figure 7.3 the vertical scale measure is 1 − IACC, so that the higher the number the stronger the spatial impression.

Figure 7.3(a) shows inverted IACC data from Ando (1977) so that it relates to "preference" on the vertical scale. This is calculated as 1 − IACC. This quantity is also the one shown for the Hidaka et al. (1997) measurements of $IACC_{E3}$ so that it relates to the ASW (image broadening) component of spaciousness. It is only important to observe that all curves start at 0 at 0°, reach a broad maximum around 60°, and then slowly decline towards 180°; differences in vertical scaling are unimportant in this context as we are looking for trends. Figure 7.3(b) shows the Hidaka et al. data reformatted as a polar plot and mirror-imaged to show left and right hemispheres. Shown in this manner the relative importance of differing reflection angles can be clearly seen. In Figure 7.3(c) the intention is made even more clear with the addition of the image of a listener. A circle of constant effect is shown, confirming that for a direct sound arriving from 0°, reflections from about 50° to 70° are likely to generate the strongest impression of ASW/image-broadening, although those from many other angles make substantial contributions.

Figure 7.3 shows a form of this measure (1 − $IACC_{E3}$) which in effect inverts the IACC vertical axis so that it matches the "preference" and "spatial impression" subjective data. Also shown are the Ando data in the same form. The results are asymmetrical front and back, because of the lack of symmetry of the human head and the placement of the ears on it. Other than a vertical scaling difference (different sounds and measurement

methods) the shapes of the curves reflect a similar trend. Because of this and the fact that the new data cover both front and back, the next step was taken.

In Figure 7.3(b) the Hidaka et al. data from (a) were plotted in polar graphical form. This gives a more intuitive expression to what seems to be a reasonable measure of "potential ASW/image broadening" for reflected sounds arriving from different directions. If the direct sound arrives from 0° it is clear that reflections arriving from about 60° will have maximal effect, but there is a wide range of lateral angles over which reflected sounds will contribute to the illusion. In contrast, sounds arriving from close to the forward and rear direction will have little effect.

All of this takes on greater meaning when in Figure 7.3(c) the shape is superimposed on the plan view image of a listener. Put this into a room plan, as we will, and it should be possible to see where reflections need to come from for maximal "preference", "spaciousness", "spatial impression" or, most precisely, "ASW/image broadening". Thinking ahead, those reflections can be the natural reflections of loudspeaker sounds in the listening room, reflections included in a multichannel recording and reproduced through one of several loudspeakers or, as will inevitably be the case, a combination of the two.

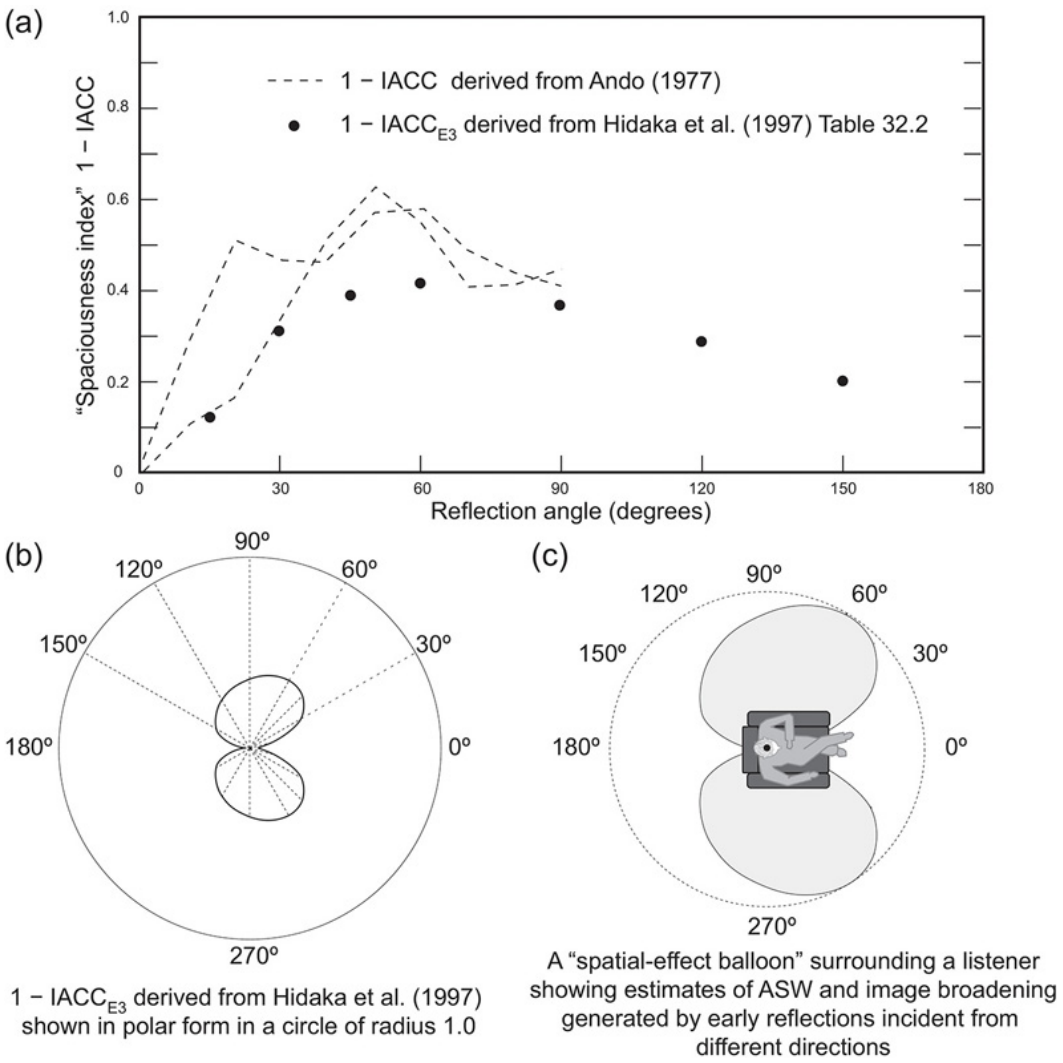


FIGURE 7.3 *Results of data from Ando (1977) and Hidaka et al. (1997) relating interaural cross-correlation (in fact: 1 − IACC), an estimate of spaciousness, to the horizontal angle from which delayed sounds arrive. The direct sound arrives from 0° and reflections arrive from angles ranging from 0° to 180°.* ⏎

It is not difficult to understand why this happens. First, because of where the ears are located on the sides of the head differences in the sounds at them will be greatest for sounds arriving from the sides, and least for sounds arriving on or close to the front-back axis. Secondly, the asymmetrical shapes of the external ears provides some acoustical gain for sounds, especially short wavelength/higher frequency sounds, arriving from

the forward hemisphere. This is inevitably linked to HRTFs but must also incorporate the manner in which the auditory processes combine the sounds at both ears. A measure of this is available in studies of directional loudness, in which listeners compare the loudness of sounds arriving from different directions. Robinson and Whittle (1960) and Sivonen and Ellermeier (2006) are good examples.

In Figure 7.4(a) the direct sounds from a conventional 5-channel surround system are shown, illustrating that the ±30° front left and right loudspeakers and the ±110° surround side loudspeakers are well located for generating pleasing spatial effects. However, for music, additional channels at the optimum angles of ±60° could be beneficial. In some current multichannel systems they appear as "front-wide" channels, but they are not always used. Dolby Atmos currently recommends them in their music production facilities.

Figure 7.4(b) shows that when the first reflections are added those from the side walls should contribute positively, but others from the front and rear walls appear to add little or nothing. If one is looking for places to add acoustical absorption to reduce overall reverberation time in a room, these would be good choices. Diffusers in those locations reduce the amplitude of the first reflections, but by sending energy off in other directions the reverberation time will be little changed.

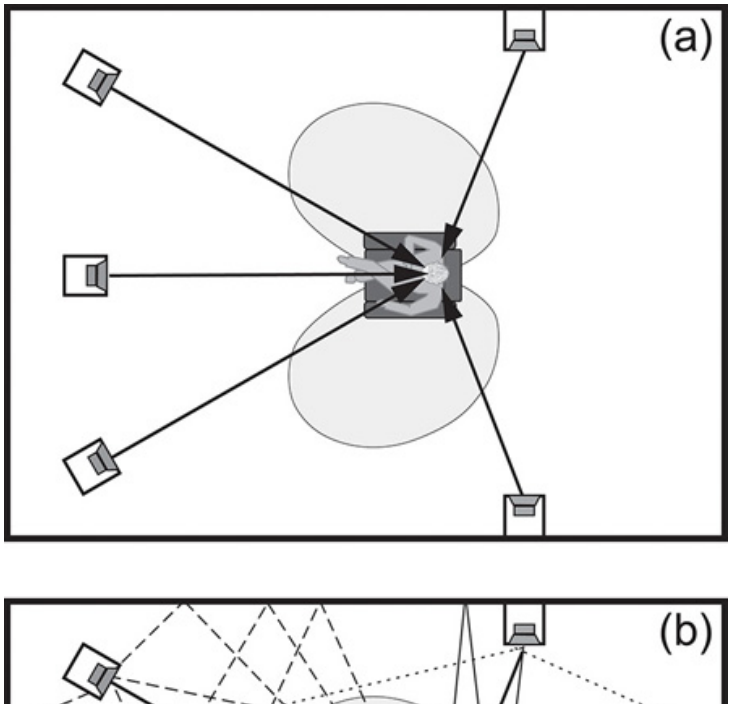
Figure 7.4(c) simplifies the presentation to a stereo system, again showing that side-wall reflections should contribute positively to spatial impressions, but those from front and rear walls do little.

The benefits being speculated on relate to natural small room reflections and do not factor in the long delays and elevated sound levels listeners indicate that they prefer, as seen in Figures 7.1 and 7.2. These are the motivation for multichannel audio, with the enhanced reflections being incorporated into recordings.

Having reached this stage in attempting to understand what might be needed for a multichannel audio system, let us take it to the next step and design one. This design is biased by my combined interest in music and movies. I enjoy multichannel music, mostly upmixed stereo, and video

concerts, many of which have multichannel soundtracks. Movies frequently have well-recorded multichannel theme music that benefits from the arrangement. I enjoy movies, but frankly am not interested in the precise localizations of off-screen sound effects, except that flyovers and similar dynamic front-to-back, back-to-front, sound effects are important in many scenes—rear channels are needed. A door slam or gunshot does not need to be precisely at 90°, 110° or whatever angle is specified for the side surround channels. Figure 7.5 shows my personal optimization for a 7-channel system. My present downsized room does not permit more channels, which would preserve all options.

Figure 7.5 shows a modified 7-channel system, designed to maximize spatial impressions for music, while preserving the essence of side channel localizations and front/back contrast and movement effects in movies. In immersive systems using 9 horizontal channels, sides can be at the usual ±90° to ±110° and front-wide channels incorporated at ±50° as has been recommended by Dolby for Atmos music applications, or ±60° as others have suggested. None of this requires laser precision in setup, so customers should feel free to experiment. In reality, some of this is determined for us by restrictions in the rooms being used. This is entertainment.



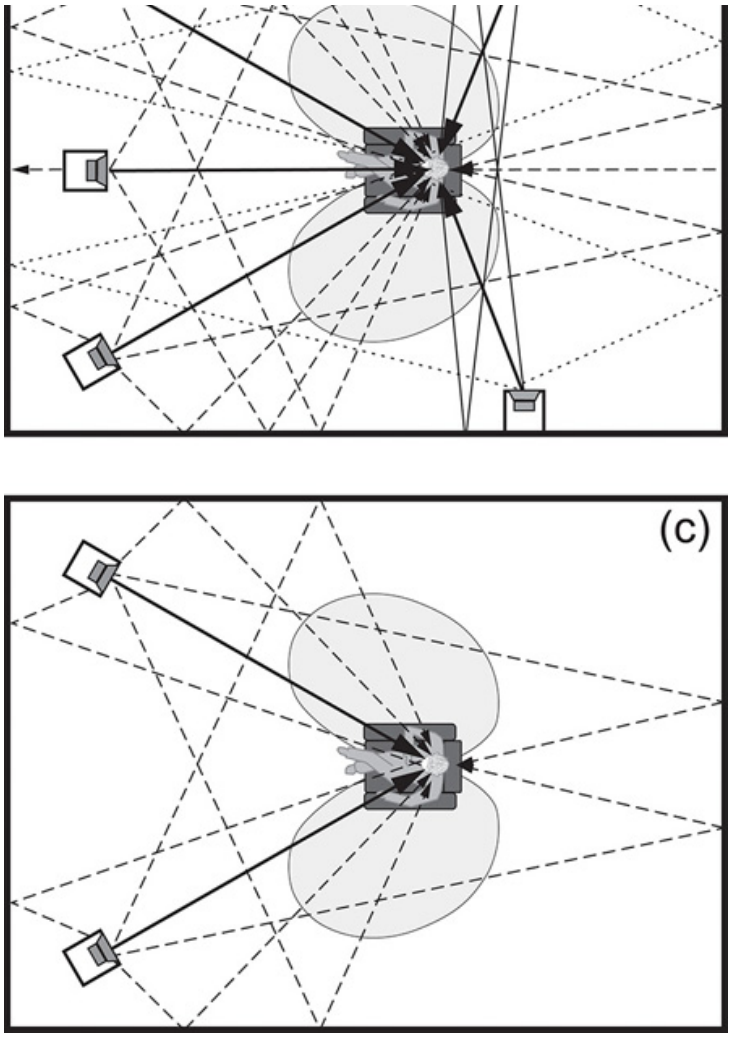


FIGURE 7.4 *revisits the acoustical ray-tracing exercise in Figure 5.12, this time superimposing the "directional preference balloon" shape from Figure 7.3(c) to see what is happening in some common listening situations.* ⏎

If the precision of horizontal angular localization is examined, we find that humans are not infallible. With speech, localization blur (uncertainty) is about 0.9° to 1.5° in the forward direction, which is good. Localization to the sides is much less precise, with an angular uncertainty range of about 20° for noise pulses, compared with about 7° in front and 11° in the rear (Blauert, 1996). Other than the front L, C, R loudspeakers, small angular modifications to locations may or may not be noticed.

7.1.3 Perceptions of Space and Sound Quality Compared

The Klippel (1990a, 1990b) papers describe an investigation that set out to relate listener descriptions of what they heard, the perceptual dimensions, to measured quantities. He summarized the process as follows:

> All dimensions perceived in the performed listening tests correspond with features extracted from the sound pressure response of the diffuse and direct field at the listener's position. There was no hint of a relation to phase response or to nonlinear distortion.

He concluded that the relevant loudspeaker measurements are the anechoic on-axis frequency response and the sound power response, the difference between which is the directivity index. This agrees with the findings described in this book, where the spinorama presents the direct sound and a sufficient collection of off-axis measurements to describe the reflected sounds arriving at the listening location—directivity index, calculated using sound power is obviously useful information.

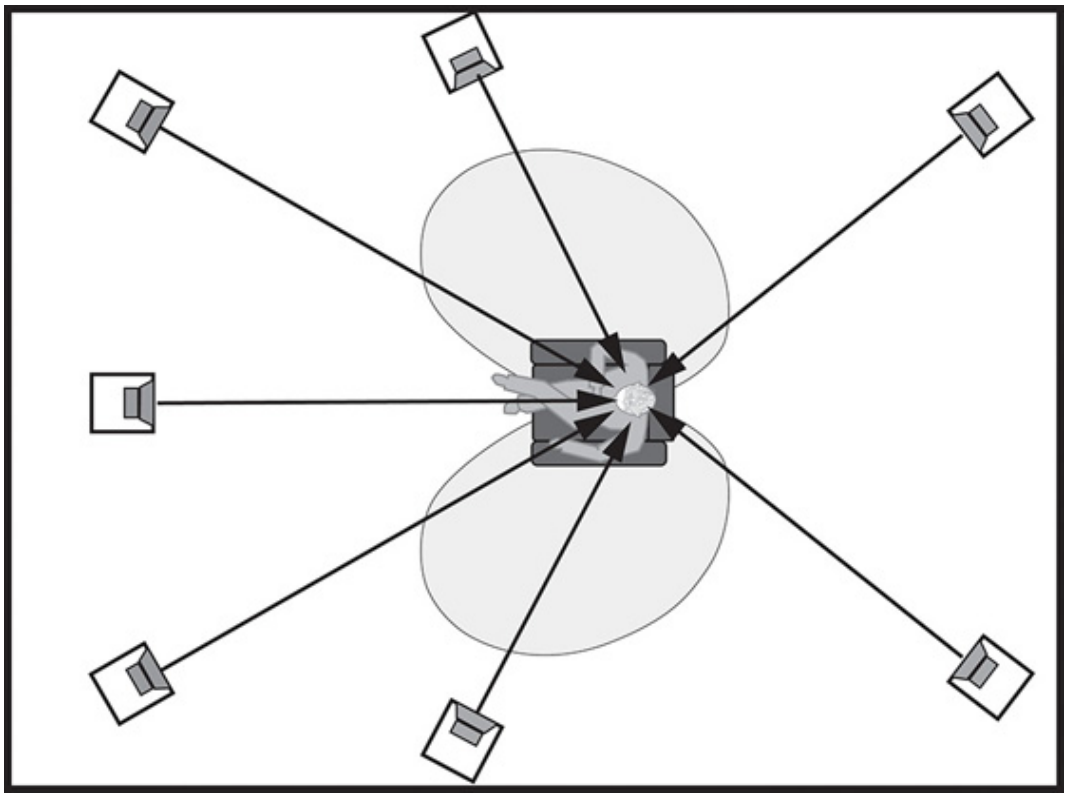


FIGURE 7.5 *This floor plan shows a 7-channel system with conventional L, C, and R locations, and conventional left and right rear locations, but with the side channels moved slightly forward to about ±65°.*

Of special interest was his finding that what he called "feeling of space" figured prominently in listener responses. In evaluating listener perceptions Klippel assessed the listener responses against what was judged to be the ideal quantity of each perceptual parameter. Thus, the perceived quality is evaluated as a defect:

$$\text{Defect} = |\text{basic measure} - \text{ideal value}|$$

A defect can therefore indicate that there is too much or too little of a perceived dimension; listeners responded according to what seemed to them to be appropriate. The responses were solicited for two broad categories, "naturalness" and "pleasantness", one relating to realism and accuracy, and the other to general satisfaction or preference, without regard to realism.

When he analyzed the factors that contributed to the perception of naturalness it was found that:

- 30% was related to inappropriate "discoloration" (sound quality).
- 20% was related to inappropriate brightness, which was explained as a 70% excess of treble and 30% lack of low frequencies (also sound quality).
- 50% to defects in the "feeling of space".

For the second general measure of quality, "pleasantness", the factor weighting was:

- 30% inappropriate discoloration (sound quality).
- 70% defects in the "feeling of space".

Thus, sensations of sound quality and spaciousness contributed equally to impressions of "naturalness" and spaciousness dominated the impression of "pleasantness". Therefore, whether one is a picky purist or a relaxed recreational listener, the impression of space is a significant factor.

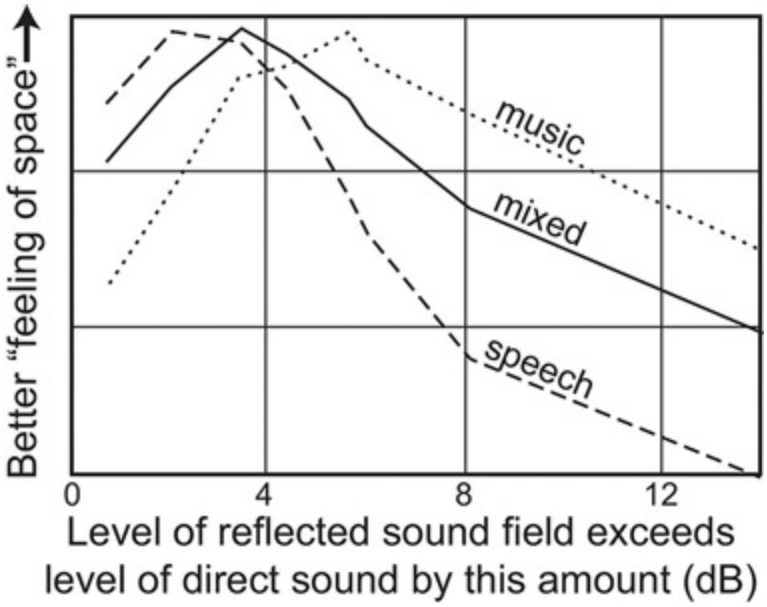


FIGURE 7.6 *The relationship between the subjective evaluation of "feeling of space" on the vertical axis, and the objective measure R (the estimated difference in sound level between the direct and reflected sounds at the listening position) on the horizontal axis. Moving up the vertical scale corresponds to increased subjective satisfaction; moving to the right corresponds to increased reflected sound in proportion to the direct sound. The maximum point of each curve describes the optimum level of reflected sound compared to the direct sound, for each of the three programs. This is an adaptation of figure 5 in Klippel (1990a).* ⏎

Perceptions of space are substantially related to laterally reflected sounds. In small rooms, it is improbable for natural reflections to initiate impressions of envelopment; those cues must be in the recordings. However, sensations of ASW (apparent source width), image broadening

and early spatial impression are very real and, it seems, desirable. Klippel chose as his objective measure of "feeling of space" (R) the difference between the sound levels of the multidirectional reflected sounds and the direct sound at the listening location.

Figure 7.6 shows that there is an optimum amount of reflected sound; there can be too much and too little depending on the nature of the program. This is in line with the evidence that the recording itself is a major factor; the amount of reflected sound in a multichannel recording is decided during the mixing operation. The smallest amount is required to provide a satisfying setting for speech, more is required for music, and music has many varieties. The optimum difference between the direct and reflected sound fields is about 3 dB for speech, 4 dB for a mixed program and 5 dB for music. There is no frequency dependence considered in these numbers, and we know that most loudspeakers do not exhibit constant directional behavior at all frequencies.

Finding a real-world interpretation of these results is not straightforward. If we are discussing single loudspeakers, it is not realistic to optimize consumer products for a single purpose, speech or music, for example. Real-world listening is usually stereo and occasionally multichannel. Voices are normally reproduced in mono (center channel), or double-mono in stereo, so it is reasonable to think that the optimum starting point for a listening room is one that adds the right amount of "space" for speech. If normal conversation is comfortable, the room is probably about right, and we know from data and experience that this is a room with measured RT in the range 0.3 to 0.5 s, in other words, typical furnished domestic rooms and home theaters. The necessary reflected sound fields necessary to deliver a pleasurable amount of "space" for music are added by the recording engineer to the stereo or multichannel recording. This is the present real-world situation.

In summary, acoustically "normal" listening rooms appear to be good starting points. Listeners preferred much higher levels of reflections than are provided by listening rooms, supporting a need for multichannel delivery systems. Preferences indicated in these tests indicate that 30% to

50% of overall satisfaction is contributed by sound quality and 50% to 70% by ambiguous “feelings of space”.

7.2 HOW MANY CHANNELS DO WE NEED FOR ENVELOPMENT?

One of the main advantages of multichannel audio is its ability to deliver a persuasive sense of envelopment—of being in a different acoustical space. This is a significant factor in movies where convincing viewers that they are inside the space they see on the screen, greatly enhancing the “suspension of disbelief”.

When one thinks of the classic example of envelopment, a symphonic concert in a good concert hall, one might think that a very large number of loudspeakers might be needed, located all around the listener. Fortunately, that is not the case. Back in Section 3.4 the idea of reproducing only the “perceptual essence” of an auditory experience was discussed: providing only the dominant cues that trigger our minds to fill in the missing pieces. If an impression of spaciousness is conveyed by a single reflection in an anechoic chamber, it should be no surprise that envelopment can be delivered by sound systems with some number of channels and loudspeakers. However, the number and locations of the loudspeakers necessary to accomplish persuasive envelopment may be surprising.

Interaural cross-correlation coefficient (IACC) has long been associated with spatial impressions, image broadening, ASW and the like—the more different the sounds at the two ears (the lower the IACC) the stronger the spatial illusion. Tohyama and Suzuki (1989) looked at a few arrangements of two and four loudspeakers, comparing measured IACCs to those found in a truly diffuse sound field—an all-but-unachievable goal in the real world, but an academic ideal. Results shown in Figure 7.7(a) indicate, not surprisingly, that 2-channel stereo did not come close to replicating the diffuse-field IACC. The real news, though, is that doubling the number of channels by adding a symmetrical pair of loudspeakers behind the listener at the same angular separation did not really change anything. The solid

dots (two channels) and open dots (four channels) in the figure are very similarly placed, and neither matches the target curve. This was the original layout for quadraphonic sound, shown in Figure 8.1, confirming that it was not an effective use of channels and loudspeakers.

When the extra pair of loudspeakers were deployed at several angles between ±45° and ±75°, as shown in Figure 7.7(b), the IACC results all moved closer to the target curve. Two things have changed: all four loudspeakers are in front of the listener, and they are at different horizontal angles. What is responsible for the improvement? From what was seen in Section 7.1.2 it is the difference in the angles that is the key—the average is ±60°, the most effective angles.

Having all loudspeakers in front of the listeners is not a problem for music, but movies often employ special effects that require loudspeakers behind listeners. One could debate the value of being able to place musicians behind listeners, but there is no debate that in movies one needs to be able to perceive flyovers and the like. Therefore, in future developments for movie applications a rearward bias was preferred, but it is worth bearing in mind that forward locations around ±60° have also have advantages for music or any movie scene that would be more persuasive if listeners felt enveloped in the sound of a different acoustical space. In that respect it is encouraging that current recommendations for Atmos music production employ front wide channels at ±50°. Sadly, few listeners in the real world will have them.

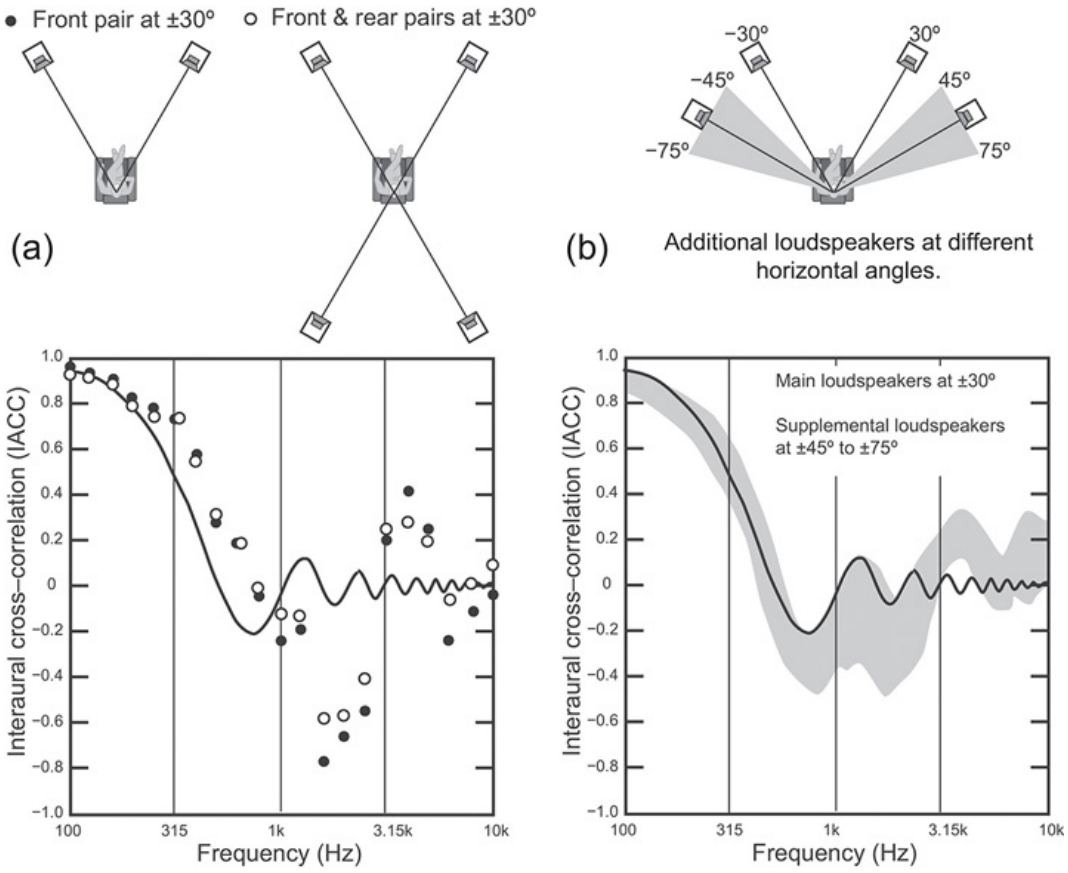


FIGURE 7.7 *(a) A comparison of IACC measured for a stereo pair of loudspeakers at ±30° and with a second pair of loudspeakers added at the same angles behind the listener, as in quadraphonics. Signals delivered by the individual loudspeakers were narrow bands of uncorrelated noise, as might be recorded in a perfectly diffuse sound field. Figure 7.7(a) shows the target curve, the solid line, the IACC of a perfectly diffuse sound field. The measured IACCs for stereo and quadraphonic configurations are similar and both deviate significantly from the target curve. Figure 7.7(b) shows results with the second loudspeaker pair moved from the rear to front locations at angles within the range ±45° to ±75°. All the measurements fall within the shaded area, which is closer to the theoretical ideal. Adapted from Tohyama and Suzuki (1989), copyright Acoustical Society of America.* ⏎

The important message in these data is that the placement of loudspeakers is as important as the number, and placement to the side is more advantageous than placement to the rear. Ironically, a 1971 publication titled “Subjective Assessment of Multichannel Reproduction”, Nakayama et al. (1971) showed that listeners preferred surround loudspeakers positioned to the sides, compared to ones placed behind, awarding subjective rating scores that were two to four times higher. It seems that nobody with any influence in the quadraphonic era read it when considering how the additional surround sound loudspeakers should be positioned.

The locations of the ears then determine that sounds arriving from different directions generate different amounts of IACC and perceived ASW. Sounds from the sides are most effective, those from front and back, least effective. It is also known from concert hall studies that diffusion in a sound field is a desirable property for distributing the sound, but that diffusion—or at least directional diversity in many reflections—is not a requirement for the perception of envelopment.

Here it is necessary to emphasize that reflections occurring in small rooms cannot *alone* generate a sense of true envelopment. Envelopment requires reflected sounds with delays longer than about 80 ms. These can only be supplied by recorded signals reproduced through multiple loudspeakers. Additional room reflections of those greatly delayed signals might enhance the impression, but the initial delay and the appropriate directions must be provided in the recorded sound and an arrangement of playback loudspeakers. Two important experiments provide important insights, one from subjective evaluations, the other from measurements. They both arrive at the same conclusion.

Hiyama et al. (2002) conducted subjective evaluations of how closely the sound of a reference sound field, generated by a circular array of 24 loudspeakers, could be approached by arrays of smaller numbers of loudspeakers. Listeners were required to judge the degree of impairment in perceived envelopment (LEV) for each of the loudspeaker configurations when compared to the 24-loudspeaker reference array at 15° intervals. The

reference array was assumed to have no impairments. All loudspeakers radiated uncorrelated noise.

The first experiment they conducted examined the performance of different numbers of loudspeakers *equally spaced* in circular arrays. The results indicated that arrays of 12, 8 and 6 loudspeakers did well at imitating the perceived envelopment of the 24-loudspeaker array. Arrangements of four and three loudspeakers did poorly, however. It was a very clear delineation, suggesting that equally spaced arrays of four and three loudspeakers should be avoided.

Figure 7.8 shows some excerpts from the results of their numerous experiments in which two to five loudspeakers in different arrangements attempted to imitate the perceived envelopment of the 24-loudspeaker reference array (score = 0). The closer to 0 is the "difference grade", the better the subjective performance of the array. There is much more data in the paper; these selections related most closely to multichannel sound reproduction options. Results for three test signals are shown: 100 Hz to 1.8 kHz noise (the frequency range over which there appears to be the strongest correlation with envelopment), and dry recordings of cello and violin to which have been added convolved and simulated early and late reflections from the appropriate directions for a concert hall. The results are interesting.

- Two-channel stereo does not fare well (a), and neither does the "quad" arrangement (f), which performed very similarly, confirming the results of Tohyama and Suzuki, discussed above.

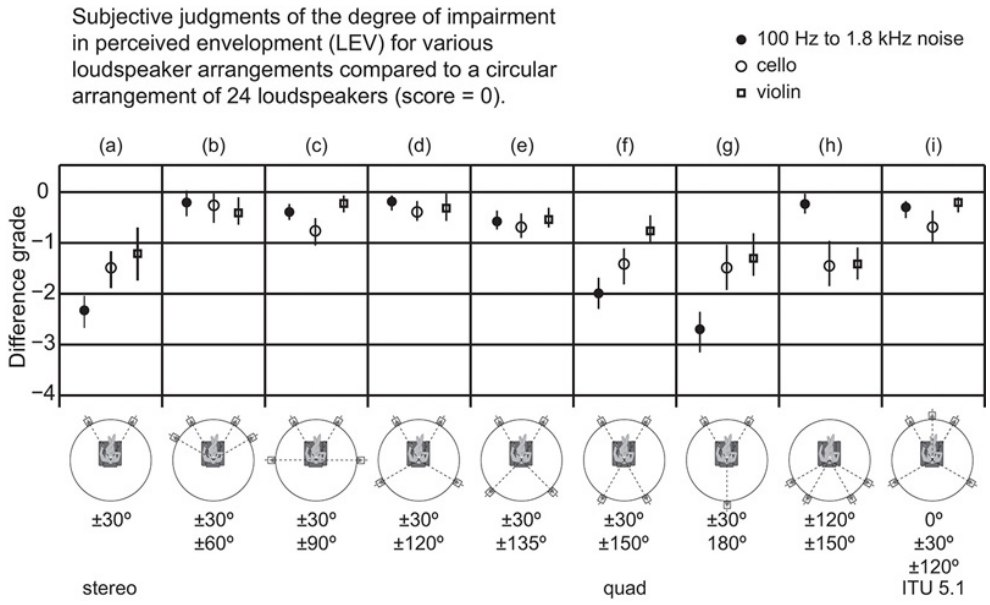


FIGURE 7.8 *Results of subjective comparisons in which listeners judged the degree of impairment in perceived envelopment (LEV) when an array of a smaller number of loudspeakers attempted to imitate the performance of a circular array of 24 loudspeakers. The data points are means with 95% confidence intervals. Adapted from the data in Hiyama et al. (2002).* ⏎

- Symmetrical front-back arrays, it seems, contribute nothing to envelopment, but add two more locations for special-effects sounds in movies and voices or instruments in music.
- A center-rear loudspeaker is worse (g).
- All combinations of a pair of loudspeakers at ±30° and another pair of loudspeakers at angles from ±60° to ±135° perform superbly (b), (c), (d), (e). Avoid ±150° (f), or whatever angle matches the spread of the front loudspeakers.
- Four loudspeakers behind the listener (h) do not perform as well as four in front, at the same reflected angles (b).

- The 5-channel arrangement described in ITU-R BS.775–4, shown in (i), performed about as well as any other configuration, but it has an added front center channel.

The last statement is obviously very good news, as it is the basis of conventional 5.1 systems.

Muraoka and Nakazato (2007) used a measurement of frequency-dependent interaural cross-correlation (FIACC) as a measure of the successful "sound-field recomposition". The idea was simple: FIACC was measured in four large spaces: a large lecture hall and three concert halls. An omnidirectional loudspeaker on the stage was the source. Recordings were made at the measurement locations using a circular array of 12 equally spaced microphones. These recordings were reproduced through different numbers of these channels using loudspeakers placed at 2 m from a measurement mannequin in an anechoic chamber. The FIACC of the sound field reproduced by each of the loudspeaker configurations was measured, and a "square error" metric was computed, describing the degree of difference between the FIACC at the original location, and that reproduced by the test arrangement of loudspeakers. The difference was computed over a "full" bandwidth (100 Hz to 20 kHz) and over a "fundamental" bandwidth (100 Hz to 1 kHz), the frequency range believed to be most related to the perception of envelopment.

Figure 7.9 shows selected data from the experiments.

This is a very different kind of experiment from that of Hiyama et al., and yet the conclusions are almost identical to those listed above. There are some minor differences, as might be expected, because here the goal was to replicate the sound field of real rooms, not a mathematical ideal, or a synthesized approximation. Some observations:

- Of special note in the top-row results is the similarity between (e), the 12-channel system, and (i), the widely used 5-channel "home theater" arrangement. *"More" is clearly not better*, at least for a single listener in the prime location.

- All of the combinations of a front pair of loudspeakers at ±30° and another pair of loudspeakers at angles from ±60° to ±120° performed reasonably well.
- The front combination of four channels (b) performed quite well, but the rear version (h) did less well, as it did in Figure 7.8(h).
- Stereo (a) performed poorly, as did the front-back symmetrical "quad" arrangement (f).
- The lower row of results shows *the significant positive effect of adding a center channel* to (b), (c), (d) and (f), creating optional 5-channel configurations, including a repetition of the ITU arrangement as (l). The already good performance of the 4-channel versions is improved, with (j), (k) and (l) all exhibiting highly attractive results.
- The front-back symmetrical arrangement (f) is slightly improved by the addition of the center channel (m), but it is still not an attractive option. The lesson: *avoid symmetrical front/back arrangements of left/right loudspeakers in 5-channel systems.*
- Again, the standard ITU arrangement of five channels performed about as well as any other configuration—good news.

This is *very* encouraging. Large numbers of channels are not necessary to provide convincing perceptual impressions of enveloping sound fields. This is true whether the evaluating method is subjective or objective. The optimal selection of four or five channels and loudspeakers can provide performance very similar to circular arrays of 12 or 24 loudspeakers. Only if one wishes discretely panned localizable images at other locations would one need more than five channels surrounding a listener, and for movies that is primarily in the rear, a solution provided by 7-channel systems.

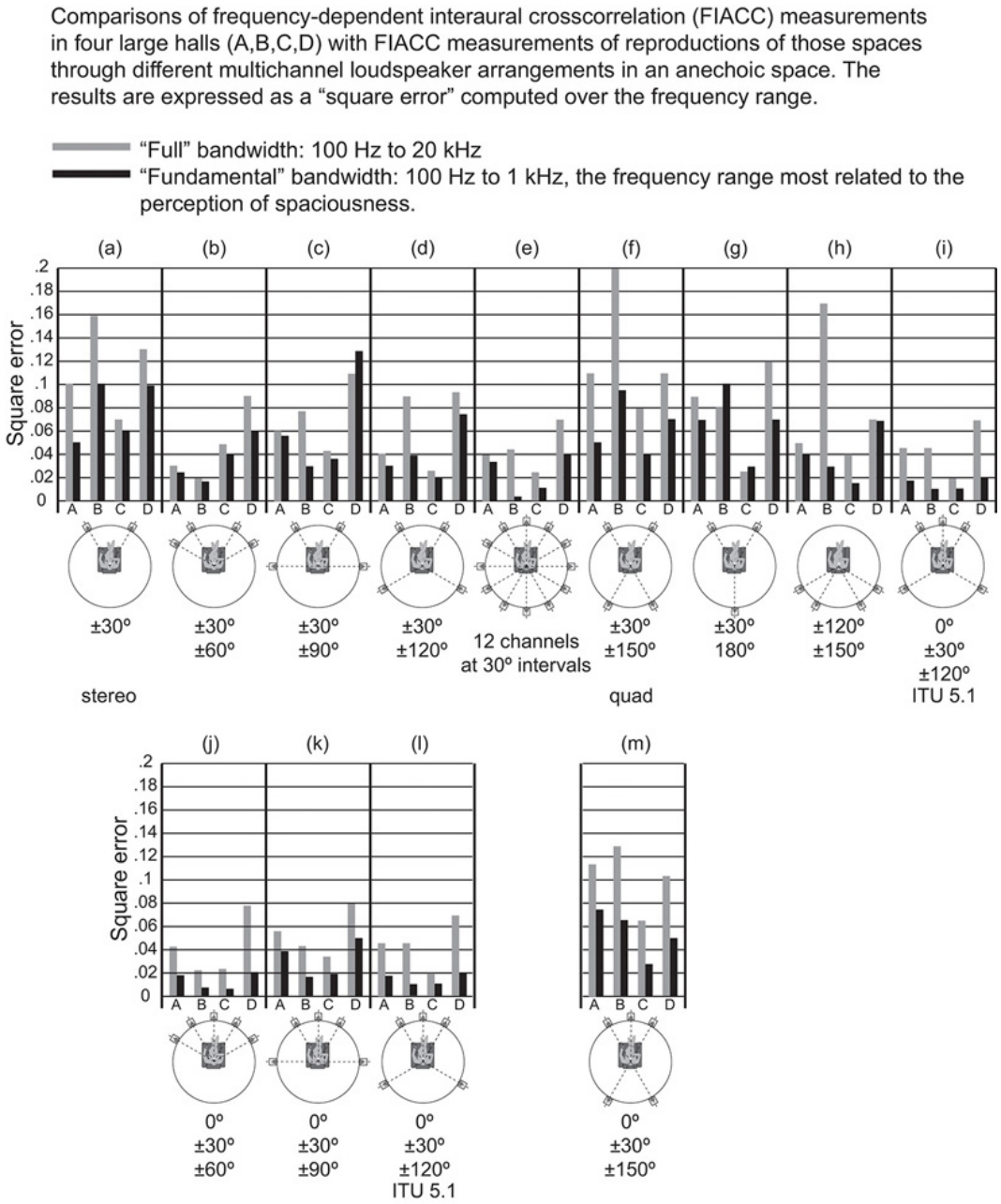


FIGURE 7.9 *Selected data from [Muraoka and Nakazato (2007)] showing how closely various configurations of reproduction channels and loudspeakers, (a) through (m), can replicate the FIACC measured in four large venues: a large lecture hall "A" and three concert halls "B", "C", and "D". The shorter the vertical bars, the better the reconstruction of the original sound field. It is probable that the black bars are more meaningful. (e)*

shows results for what should be the best possible configuration: 12 channels at 30° intervals. ⏎

NOTE: The previous sentence focused on "a listener". If there is an audience of multiple listeners located away from the symmetrical sweet spot, more channels will be necessary to generate similarly enveloping effects over the enlarged listening area. This is where the more elaborate immersive systems with added elevation channels have advantages. In the demonstrations I have heard it was impressive how stable the spatial illusions were as I moved around the listening room. More channels clearly can deliver a more believable sound field for multiple listeners.

All these experiments were conducted in anechoic listening circumstances. Reflections within the listening room will have some effect on the conclusions, but it is highly probable that the direct sounds from the loudspeakers will have the dominant effects and, the more channels, the greater that dominance is likely to be. Hughes et al. (2016) offers evidence that the larger of two sound fields in a room is the one that listeners tend to hear, so reflections within the listening room may not be a major problem.

7.3 ELEVATION: THE MISSING DIMENSION

More than a century of listening to reproduced sound in two dimensions has been very rewarding. We have a good understanding of the properties of binaural hearing and sound localization as applied to horizontal plane arrangements of loudspeakers. Now, we have height channels and overhead channels to augment spatial presentations.

Because human ears are in the horizontal plane we can localize with great precision in that plane. Differences in the timing (interaural time difference, ITD) and amplitude (interaural amplitude difference, IAD) of sounds at the two ears provide us with the basic binaural cues. This is well-explored territory, and I won't go into it. However, localization in the vertical plane, in elevation, is different. Here the ITD and IAD cues still contribute information, but only when the head is in motion and the sound

source is stationary. The amount of movement to demonstrate the effect can be substantial—e.g. ±30° swings in the Perrett and Noble (1997) experiments, a ridiculous amount not employed in recreational listening. For home and cinema entertainment head movements are unlikely to contribute much information, and the sound sources of interest are not always stationary, especially in blockbuster movies with aircraft and superheroes whizzing overhead. Obviously, there must be more information.

In the important median plane we rely on spectral cues derived from the way sounds interact with our external auditory apparatus: the pinnae. When I was at the National Research Council of Canada a colleague, Dr. Edgar A.G. Shaw, was investigating the acoustics of the external ear, exposing the acoustical reasons why sounds arriving from different angles had different timbral signatures at the eardrums. These were the precursors to what we now call head-related transfer functions: HRTFs. The mechanisms include acoustical interference and resonances in the folds of the pinnae and the cavities within them. Several of them are sensitive to the direction of the incoming sound, exhibiting different levels of response for different directions.

Looking at ears, the differences we can see among subjects result in very different acoustical performances (Shaw, 1974a, 1975, 1982): peaks and dips of ever-changing shapes. Some researchers have been paying attention to the highly visible acoustical interference dips in the curves. Such dips are highly variable among subjects, but in real listening situations the direct sound does not exist in isolation. There are numerous reflected sounds arriving from many directions that confound any interference dip cues. None of this is stable, changing with even slight head movements. It is difficult to imagine why dips would be the focus of natural selection in the evolutionary process.

It is important to realize that the localization cue we are talking about is really perceived timbre, and it is known from sound quality research discussed earlier in the book that listeners are very sensitive to resonant peaks and relatively insensitive to dips—acoustical interference dips are

largely ignored. Underlying the untidy measured frequency-response curves is evidence of relatively stable resonances. Humans pay attention to, and are very sensitive to resonances, as they are significant timbral cues for all the sounds we are interested in: voices, musical instruments, etc. (Sections 1.5.1 and 5.3, and Toole and Olive, 1988).

Figure 7.10 summarizes the conclusion that Shaw came to, showing that averaged across several subjects, there was a progressive increase in the energy in a resonance around 7 to 8 kHz and a decrease at frequencies around 12 kHz. The change is dramatic.

These are historical data pertaining to external ear acoustics, and not complete far-field measured HRTFs. Modern measurements of this are needed. However, what is reported here is a component in the overall HRTF, and this evidence indicates that to be able to localize sounds in elevation:

- The sound must contain frequencies above about 6 kHz, and listeners must be able to hear it.
- Wide bandwidth and a dense spectrum help to reveal the height cue.
- Familiarity with the sound is important to recognize what the spectrum means. It is a learned skill which in situ involuntary head movements very likely assist in.

We are able to localize sounds in the horizontal plane with good accuracy because of the powerful interaural time and amplitude differences. With speech localization blur (uncertainty) is about 0.9° to 1.5° in the forward direction. As mentioned earlier, localization to the sides is much less precise, with an angular uncertainty range of about 20° for noise pulses, compared with about 7° in front and 11° in the rear (Blauert, 1996).

However, vertical localization is mainly a spectral/timbral recognition exercise, which is a very different phenomenon. Vertical localization “blur” (uncertainty) is about 17° for an unfamiliar voice, about 9° for a familiar voice and 4° for white noise (Blauert, 1996). Being more precise about

vertical angles, the range of angular uncertainty for continuous speech by a familiar person is about 18° horizontally in front, 20° at 30° elevation, 26° at 74° elevation, and 44° at 18° behind directly overhead (Damaske and Wagener, 1969, as reported in Blauert, 1996). These are huge uncertainties, however confident we feel about our perceived localizations. If we have no idea of the spectral content of a specific sound, there is no assurance that we would automatically perceive it at its true elevation. Does the precise location of a real or phantom image in the overhead immersive imagery truly matter if none of the sounds up there is familiar to us?

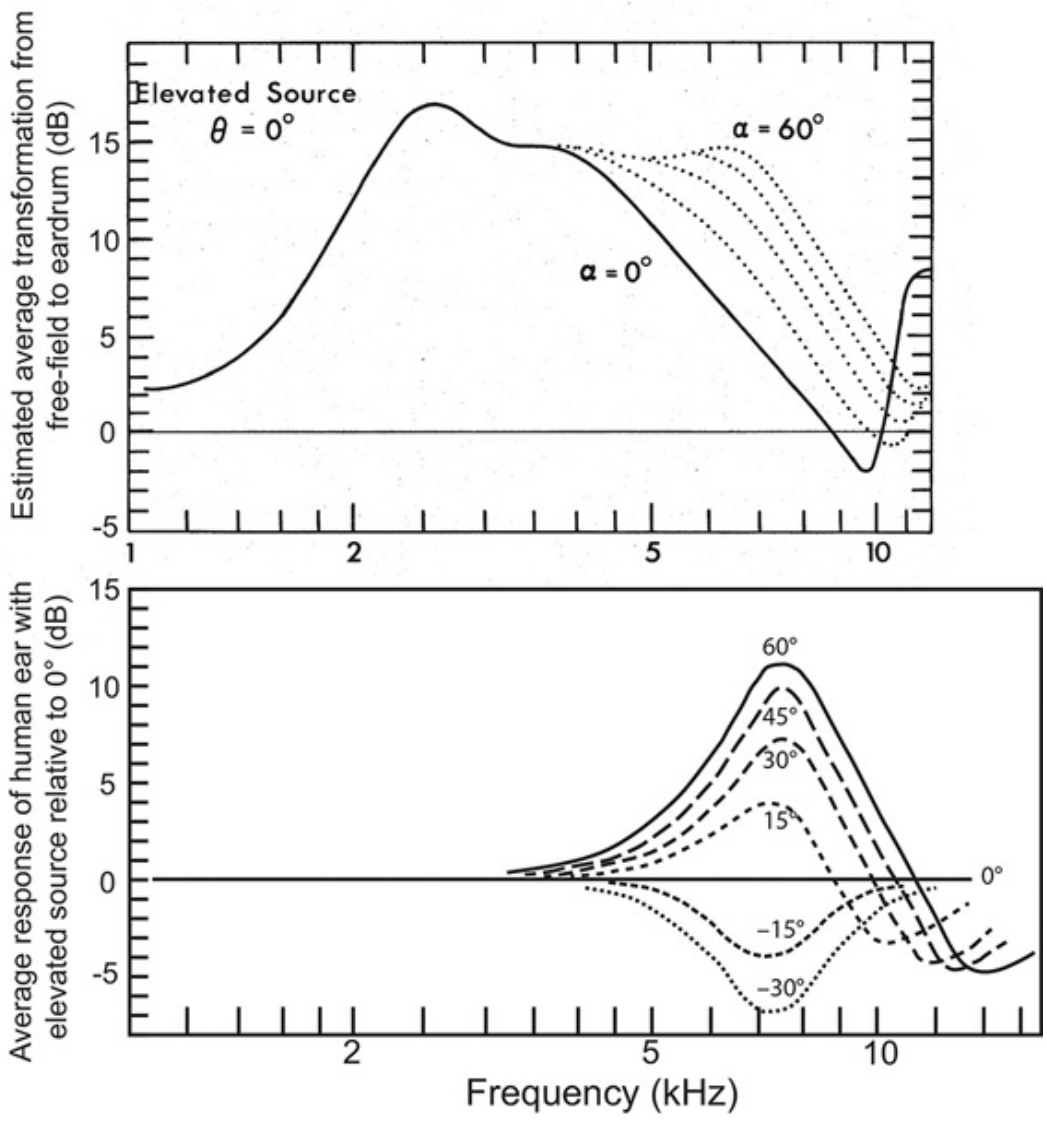


FIGURE 7.10 *(top) The mean curve from ten subjects shows a progressive increase in energy around 7 kHz as a function of source elevation angle. Data from Shaw, (1974b), copyright Acoustical Society of America. The lower curves show evidence of energy in the 7 kHz resonance being elevated and another around 12 kHz decreasing as a function of source elevation. Data from Shaw, personal correspondence, September 29, 1972.* ⏎

We rely greatly on "plausibility" when localizing in complex situations. Aircraft are automatically assumed to be above us, even in traditional 5.1 and 7.1 horizontal plane multichannel systems. Nothing is localized below the floor. Voices are where the lips are on the screen—the ventriloquism effect. If listeners are predisposed to think that sounds are coming from above, that is what they are likely to be perceived, hence the emphasis on hovering helicopters and bird sounds in demonstrations of immersive sound systems.

It is fortunate that the elevation cues are at high frequencies. It means that bouncing some high frequencies off the ceiling can assist the illusion, as has been shown with small loudspeakers aimed upwards in Atmos systems. The effect is not as persuasive and the localization not as durable as with loudspeakers in ceiling locations, but such installations are not always possible or practical. The ceiling-located loudspeakers must be able to deliver the necessary high frequencies to all listeners, which is a factor to consider in installations—off-axis output degrades rapidly at frequencies above 6 kHz.

These directional cues are built into each of us and we are individualistic, so learning the elevation cues in situ is part of the purpose of head movements. Whether we attend to resonant peaks, interference dips, or both, variations among listeners are likely to be permanent features.

REFERENCES

Ando, Y. (1977). “Subjective Preference in Relation to Objective Parameters of Music Sound Fields with a Single Echo”, *J. Acoust. Soc. Am.*, 62, pp. 1436–1441. ⏎

Ando, Y. (1985). *Concert Hall Acoustics*, Springer-Verlag, Berlin. ⏎

Ando, Y. (1998). *Architectural Acoustics*, Springer-Verlag, New York. ⏎

Blauert, J. (1996). *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, MA. ⏎

Cremer, L. and Müller, H. A. (translation by T. J. Schultz) (1982). *Principles and Applications of Room Acoustics*, Vol. 1 & 2, Applied Science Publishers, London. ⏎

Devantier, A. (2002). “Characterizing the Amplitude Response of Loudspeaker Systems”, *113th Convention, Audio Eng. Soc.*, Preprint 5638.

Griesinger, D. (1989). “Theory and Design of a Digital Audio Signal Processor for Home Use”, *J. Audio Eng. Soc.*, vol. 37, pp. 40–50. See also: Lexicon CP-1 owner’s manual, available for download at www.lexicon.com ⏎

Griesinger, D. (1997). “The Psychoacoustics of Apparent Source Width, Spaciousness and Envelopment in Performance Spaces”, *Acta Acustica*, vol. 83, pp. 721–731. ⏎

Griesinger, D. (1998). “General Overview of Spatial Impression, Envelopment, Localization, and Externalization”, *15th International Conference, Audio Eng. Soc.* ⏎

Griesinger, D. (1999). “Objective Measures of Spaciousness and Envelopment”, *16th International Conference, Audio Eng. Soc.* ⏎

Griesinger, D. (2001). “The Psychoacoustics of Listening Area, Depth and Envelopment in Surround Recordings, and Their Relationship to Microphone Technique”, *19th International Conference, Audio Eng. Soc.* ⏎

Hidaka, T., Beranek, L. L. and Okano, T. (1997). “Some Considerations of Interaural Cross Correlation and Lateral Fraction as Measures of Spaciousness in Concert Halls”, Chapter 32 in Ando, Y. and Noson, D., editors, *Music and Concert Hall Acoustics*, Academic Press, London. ⏎

Hiyama, K., Komiyama, S. and Hamasaki, K. (2002). "The Minimum Number of Loudspeakers and Its Arrangement for Reproducing the Spatial Impression of Diffuse Sound Field", *113th Convention, Audio Eng. Soc*., Preprint 5674. ⏎

Hughes, R., Cox, T., Shirley, B. and Power, P. (2016). "The Room-in-room Effect and Its Influence on Perceived Room Size in Spatial Audio Reproduction", *141st Convention Audio Eng. Soc*., Preprint 9621. ⏎

Klippel, W. (1990a). "Multidimensional Relationship between Subjective Listening Impression and Objective Loudspeaker Parameters", *Acustica*, vol. 70, pp. 45–54. ⏎

Klippel, W. (1990b). "Assessing the Subjectively Perceived Loudspeaker Quality on the Basis of Objective Parameters", *88th Convention, Audio Eng. Soc*., Preprint 2929. ⏎

Muraoka, T. and Nakazato, T. (2007). "Examination of Multichannel Sound Field Recomposition Utilizing Frequency Dependent Interaural Cross Correlation (FIACC)", *J. Audio Eng. Soc*., vol. 55, pp. 236–256. ⏎

Nakayama, T., Miura, T., Kosaka, O., Okamoto, M. and Shiga, T. (1971). "Subjective Assessment of Multichannel Reproduction", *J. Audio Eng. Soc*., vol. 19, pp. 744–751. ⏎

Perrett, S. and Noble, W. (1997). "The Effect of Head Rotations on Vertical Plane Sound Localization", *J. Acoust. Soc. Am*., vol. 102, pp. 2325–2332. ⏎

Robinson, D. W. and Whittle, L. S. (1960). "The Loudness of Directional Sound Fields", *Acustica*, vol. 10, pp. 74–80. ⏎

Shaw, E. A. G. (1974a). "Physical Models of the External Ear", *Proc. 8th Int. Cong. Acoust*., London, vol. 1, p. 206. ⏎

Shaw, E. A. G. (1974b). "Transformation of Sound Pressure Level from the Free Field to the Eardrum", *J. Acoust. Soc. Am*., vol. 56, p. 1848. ⏎

Shaw, E. A. G. (1975). "The External Ear: New Knowledge", presented at the VII Danavox Symposium, Denmark. ⏎

Shaw, E. A. G. (1982). "External Ear Response and Sound Localization", Chapter 2 in Gatehouse, R., editor, *Localization of Sound*, Amphora Press. ⏎

Sivonen, V. P. and Ellermeier, W. (2006). “Directional Loudness in an Anechoic Sound Field, Head Related Transfer Functions, and Binaural Summation”, *J. Acoust. Soc. Am*., vol. 119, pp. 2965–2980. ⏎

Tohyama, M. and Suzuki, A. (1989). “Interaural Cross-correlation Coefficients in Stereo-reproduced Sound Fields”, *Jour. Acoust. Soc. Am*., vol. 85, pp. 780–786. ⏎

Toole, F. E. and Olive, S. E. (1988). “The Modification of Timbre by Resonances: Perception and Measurement”, *J. Audio Eng. Soc*., vol. 36, pp. 122–142. ⏎

CHAPTE
R 8

Audio Formats: Part Two—Multichannel Audio

Floyd Toole

DOI: 10.4324/9781003477495-9

As much as audio purists have appreciated the virtues of accurately reproduced music, soundstage, imaging, and envelopment, music has not been the driving force behind audio formats. It has been movies. Just as Blumlein was motivated by movies to investigate, and indeed invent, aspects of stereo—the minimalist multichannel format—it was movies that drove the channel count upwards. In addition to expanding the audio illusions beyond the on-screen action, multichannel systems were very persuasive in acoustically "transporting" viewers into the spaces that were seen on the screen. This is an interesting, multifaceted story (https://en.wikipedia.org/wiki/Surround_sound [accessed June 2024]).

8.1 QUADRAPHONICS: STEREO TIMES TWO—A FLAWED IDEA

For domestic entertainment the first major development after stereo was quadraphonics, employing four loudspeakers symmetrically arranged (left and right front, and left and right rear) around a listener. Cumbersome open-reel tape versions existed earlier, but it became popular when the LP version was introduced in the 1970s. Figure 8.1 illustrates the layout.

There were two categories of quadraphonic systems in use: matrixed and discrete. The encode/decode matrixed systems downmixed four signals into the bandwidth normally used for two channels. In doing this, something has to be compromised and, as a result, all channels did not have equal channel separation from the others. Information that was supposed to be only in one channel would appear in smaller quantities in some or all of the other channels. The result of this "crosstalk" was confusion about where the sound was coming from and an extreme sensitivity to listener position; leaning left, right, forward, or back caused the sound panorama to be biased in that direction. The sweet spot was very small. Spacious classical music was a favorite in demonstrations, but the ambiguity in localization was not always flattering to popular music.

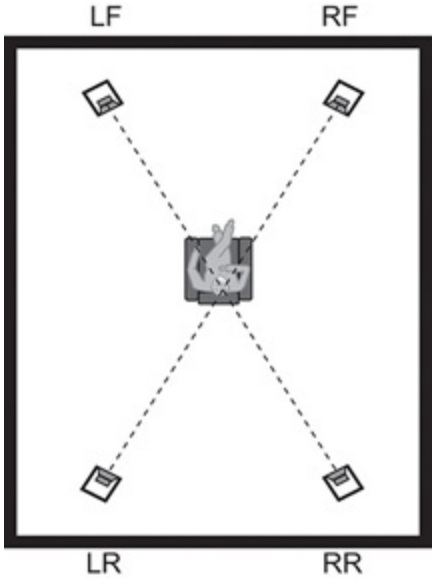


FIGURE 8.1 *A quadraphonic listening arrangement.*

Various forms of signal-adaptive “steering” were devised to deliver more convincing directional illusions during the playback process. There was SQ from CBS, QS from Sansui, E-V from ElectroVoice and others. Peter Scheiber figures prominently as a pioneer in matrix design, with his patented encoder and decoder ideas being incorporated into many of the systems. The best of these systems were remarkably good in creating the illusion of completely separate, or discrete, channels when a single image was panned around the room. However, this clear separation breaks down when there is a demand for several simultaneously occurring discrete images. In the limit, the steering ceases and we listen through the raw matrix, with its generous crosstalk leakage.

Ultimately, four discrete channels were needed. However, achieving this on vinyl LPs required that the recorded bandwidth be extended to about 50 kHz—a challenge. Nevertheless, it was accomplished, as CD-4 from JVC and, although this quadraphonic format was short-lived, the technology necessary to achieve the expanded bandwidth had a lasting benefit on the quality of conventional 2-channel LPs. Among these were half-speed cutting processes, better pressings and playback cartridges with high compliance, low-moving mass and new stylus shapes combined to yield wider bandwidth and reduced tracing and tracking distortions.

Discrete multichannel tape recordings were available, but open reel tape was an expensive nuisance, and packaged tape formats (e.g. cassettes) were not ready for true high fidelity multichannel sound. Years passed, and the industry was unable to agree on a single standard, which was an intolerable situation from a business perspective. There were issues with mono, stereo and broadcast compatibility (Crompton, 1974). Eventually, the whole thing dissolved (Torick, 1998). The industry lost money and credibility, and customers were justifiably disconcerted after many of them invested in soon-to-be-useless hardware.

Looking back on this unfortunate episode in the history of audio, one can see another reason for failure—the system was not psychoacoustically well founded. Lacking an underlying encode/decode rationale, the problems of 2-channel stereo were simply compounded. There were even notions of

“panning” images front to back using conventional amplitude-panning techniques, something that Ratliff (1974) and others found problems with—the ears are in the wrong locations for that to work. The quadraphonic square array of left and right, front and rear, was still an antisocial system, with the sweet spot now constrained in the front-back as well as the left-right direction.

One of the prime objectives of multichannel audio is to deliver envelopment, the impression of being in a different acoustical space. As was shown in Section 7.2, adding two channels in a symmetrical front/back array, as was done here, has very little positive effect. A different arrangement is needed, with a center channel for movies.

8.1.1 Dolby Adapts the Technology to a Better Purpose

All was not lost, though. The active matrixes developed during the quadraphonic era were inspired technologies that improved the directional impressions of matrix encoded multichannel programs. Dolby smartly realized the value in the technology, worked with the Sansui QS matrix for a period and then created one of their own, configured into Left, Center, Right and Surround channels, a much more effective arrangement for delivering the desired illusions, with a center channel to stabilize the soundstage. These “Dolby Stereo” soundtracks were quickly adopted by the movie industry. The “surround” illusion was created by a single monophonic channel which, in cinemas, energized multiple loudspeakers placed at intervals along the side walls and across the rear wall. It worked quite well in the large venues; see Figure 8.2.

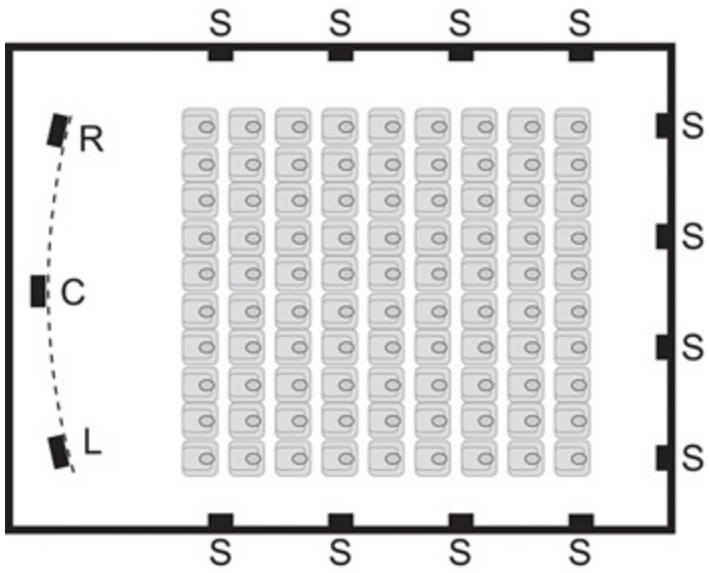


FIGURE 8.2 *Dolby stereo as it was developed in 1976: three channels across the front and a single surround channel delivered identically to loudspeakers surrounding the audience. This is the foundation from which the home theater and multichannel music technologies we know today evolved.* ⏎

For home theaters an adapted process was called “Dolby Surround” and later, when economical chip versions were possible, “Dolby Pro-Logic”. However, in small domestic home theaters the surround illusion was less than persuasive with the customary pair of side loudspeakers. In the beginning the surround signal was monophonic, meaning that listeners on the center line localized the sound in their heads, which was not acceptable. The first remedy was to reverse the polarity of one of the surround loudspeakers, which helped. A substantial improvement, introduced by Shure HTS, was to electronically decorrelate the sounds emerging from the left and right loudspeakers (Julstrom, 1987). This was very effective, and the idea, although not the Shure patented algorithm, was adopted by THX. THX approved the use of other algorithms in their licensed products, though.

THX also introduced bidirectional out-of-phase loudspeakers (incorrectly called dipoles) to attenuate the direct sound and add additional reflections to

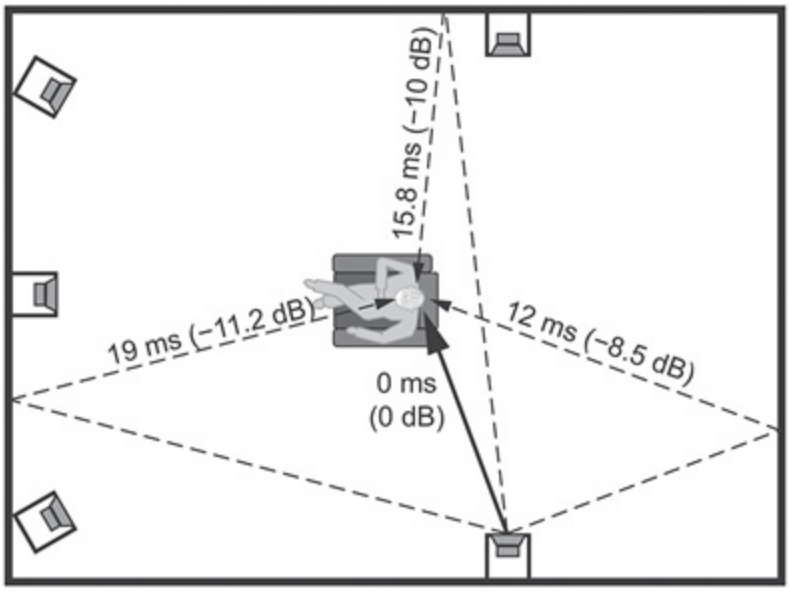
the surround sound field, claiming that the added “diffusion” would be advantageous. Diffusion is a property of a sound field, not a perception, so there was some marketing bias here. Small listening rooms are too acoustically “dead” to support a diffuse sound field. The real problem with this innovation was that the chosen design of these loudspeakers seriously degraded the sound quality of the important direct sound, while the added reflections came from directions ineffective at producing the interaural differences that yield the desired spatial perceptions. This aspect of audio history is reviewed in OR Chapter 8.

All of this was a transitional phase in multichannel developments.

8.2 DISCRETE MULTICHANNEL ARRIVES

Fortunately, it was not long before digital technology permitted five or more completely discrete channels. Recording engineers could then control the kind and strength of off-screen sounds and spatial effects. The sound delivery problem went away, and conventional forward-firing and loudspeaker configurations were recommended for use everywhere with on-wall bipole (bidirectional in-phase) as a wide dispersion option for surrounds.

(a) Discrete sounds panned to a side surround channel. *The perception: sound coming from that loudspeaker.*



(b) Surround channels provide delayed "reflections" following the direct sound from any or all front channels. Delays are as shown in (a) plus delays introduced in the recording process between the front channels and each surround. *The perception: spaciousness, envelopment.*

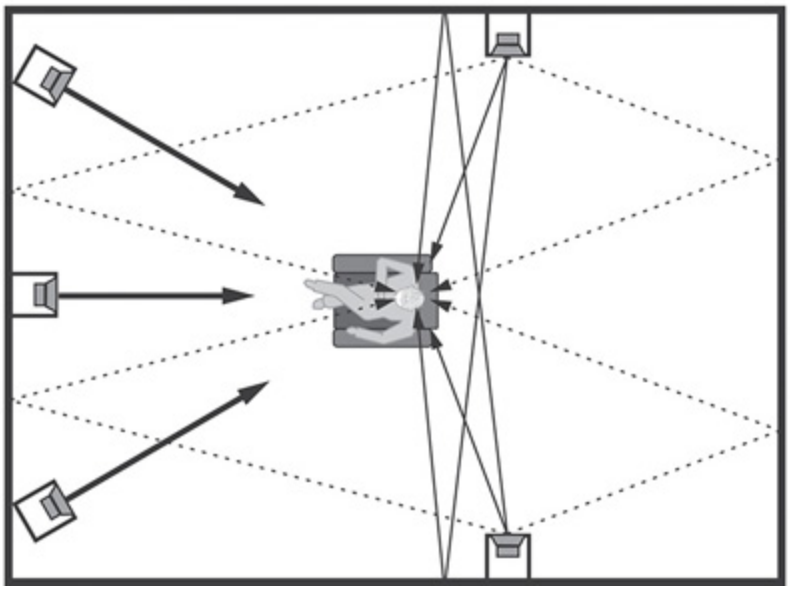


FIGURE 8.3 *Reflected sounds for two usages of surround loudspeakers: (a) as an independent sound source and (b) as a supplement to the front loudspeakers reproducing delayed sounds intended to yield impressions of distance, spaciousness or*

envelopment. Delays and sound levels shown are relative to the direct sound, assuming perfect reflections. For a 21.5 ft (6.5 m) × 16 ft (4.9 m) room. ⏎

There was a scramble to be first and best in implementing digital audio in films, a major complication being how to combine multichannel digital audio with optical film visuals. There is a complex history that will be bypassed here; it is sufficient to say that in 2024 there are few limits to audio quality or capacity in professional and state-of-the-art domestic delivery systems. Streaming services are a different story, but this is still an evolving technology.

Movies and music both benefited from the perceptions this system could generate, ranging from clearly localized off-screen sounds to appropriately delayed reflected sounds that generated persuasive envelopment. It was a good beginning. Figure 8.3 illustrates the basic effects.

8.2.1 The Basic Functions Required of Multichannel Systems

Let us begin by defining the duties of a surround-sound system. It will be assumed that such a system is intended to cater to more than one listener, although it is likely to work best for the prime listening location, where it is calibrated.

- *Localization*: The perception of direction: where the sound is coming from. The minimum number of locations would be the number of discrete or steered channels/loudspeakers in the system. Beyond that, we would need to rely on phantom images floating between pairs or multiples of loudspeakers. Perceptually the location of a sound source is defined by the first arrival, the direct sound, at the two ears.
- *Localization in a channel-based system*: In this situation, the number of sound image locations is defined by the number of discrete channels and loudspeakers. The advantage of this system is that every listener in the room localizes the sound to the same location. In movie

soundtracks it is common for some basic components of the mix to be delivered in this fashion throughout the film to provide a stable base, and to simplify the soundtrack to make edits easier. In many films much of the on-screen sound is delivered by the center channel, relying on the visual cues to direct the perception—the ventriloquism effect. Background music is similarly simplified. Music recordings most commonly are based on a stereo front soundstage that is elaborated on in various ways using the other channels.

- *Localization in an object-based system*: In action movies the limited number of real loudspeakers in practical domestic systems is a restriction to channel-based image steering, so phantom images (objects) floating between and among groups of loudspeakers are used to fill the gaps. However, as we know from stereo, the perceived location of phantom images changes with listener position. Consequently, especially in systems with few loudspeakers, the phantom image locations are reliable only for a single listener in the prime/reference location. The locations of the ears means that horizontal angle localization works well, but up-down and front-back panning is relatively imprecise. Anyone seated away from the sweet spot would hear a distorted panorama of phantom sound images, although this may not degrade the entertainment, as the intended location of off-screen sounds is not normally known, and precision is not a requirement. Conclusion: more discrete channels driving more loudspeakers are desirable. A movie mixed in a 34-channel dubbing stage and played back in a home system with the “normal” 5 to 11 channels will experience spatial changes. A fact not always acknowledged is that phantom object-based images suffer from sound quality/timbral variations as different combinations of loudspeakers are used to steer images to different locations—acoustical interference/comb filtering. It is fortunate that object-based images are frequently moving or transient, rather than stable; the flaws are less noticeable.

- *Distance*: By the appropriate addition of delayed, reflected, sounds in the recordings it is possible to create impressions of distance, moving the perceived locations well beyond the loudspeakers themselves. This impression is aided by envelopment, the impression of being in a large room.
- *Envelopment*: The perception of being in a different space, usually large, surrounded by ambiguously localized sound. This is a very important function. It requires a direct sound, as from an orchestra or band, usually at the front, followed by significantly delayed sounds, from the sides and rear, simulating reflections in the performance venue. These can be included in live recordings or added later using spatial synthesizers. Both methods can give listeners the impression of being in a different space. Convincing perceptions require low interaural cross-correlation (IACC), i.e. different sounds in both ears. This is best achieved with sounds arriving equally from both sides, something that normally only happens for listeners in the sweet spot and on the center line of a room. Certain loudspeaker designs can improve circumstances for listeners to the sides of center, as can some acoustical treatment options.
- *Immersion*: The latest evolutionary stage of multichannel. Immersive systems deliver sounds from both horizontal and elevated locations. There is ambiguity in what is meant by the term because immersion shares some properties with envelopment. In my opinion, the distinction is that immersion is the impression of being surrounded by, immersed in, sounds originating in many locations, horizontally and vertically. There is no deliberate attempt to create the illusion of being in a physical room or hall. One could be outdoors in a forest with fluttering leaves and birds all around, in the midst of a thunderstorm with rain, or surrounded by a crowd of people shouting, clapping and cheering. Systems capable of these illusions can also deliver specifically localizable images in many locations. Some definitions extend beyond auditory considerations and allow for “emotional”

immersion in the total experience, including the plot, which may or may not require an immersive sound system (Agrawal et al., 2019).

- *Upmixing and downmixing* from the number and locations of loudspeakers used to create the program to the number and locations of loudspeakers used for playback. Upmixing began with stereo recordings being processed to play through the early 5- and 7-channel versions of multichannel. These were empirical, trial-and-error operations and the results varied. Current immersive systems present a different challenge because loudspeaker numbers and locations vary at both the recording and playback ends of the process. A solution is enabled by metadata—information about the recording that is included with the recording. At playback the metadata about the recording system layout is used by a processor to modify the signals for reproduction through what might be a very different system layout. Obviously, success depends on how the system interprets information about the number and locations of loudspeakers in the playback system. Home systems are likely to be simpler than those used for program creation. These days, sound bars are proliferating, simplifying things even more in some ways, but complicating things in other ways when they attempt to use reflected sounds and signal processing as substitutes for real loudspeakers in different locations. Music is often downmixed from many immersive channels to stereo by algorithms. Ironically, many immersive mixes are created from stereo originals by recording engineers—they are artistic exercises done for commercial reasons and not all are persuasive "improvements" on the originals.

The possibilities for enhancing image localization are limited by the practical number of channels the industry feels customers will buy and install in their homes. For most people it seems that 5.1 is difficult to accommodate in their living spaces. However, for enthusiasts with the space and budget there are immersive systems up to 34 channels available for home installations. These can deliver superheroes and aliens flying around the room and overhead, with as much immersion and envelopment as one

can tolerate. A high channel count has the advantage of satisfying a large audience, which is a consideration in home theaters.

Even though the number of screen channels remains at three for most systems, for most of the time in movies all on-screen sounds are delivered by the center channel. I get up from my chair from time to time just to satisfy my curiosity because for sounds associated with on-screen action there is the powerful “ventriloquism” effect. We hear the sound coming from where we see the lips move, the gun flash, or the door slam. The hours spent by moviegoers listening to a monophonic center channel as action moves around on the screen suggests that it works very well. Only big budget movies do much panning of sounds across the screen. Why? It costs more to do and complicates the editing process that can sometimes go on up to hours before a film is released. When the locations of sounds are few and stable, editing becomes mostly a visual continuity and storytelling operation.

Sounds originating off-screen are frequently brief sound effects for which no real precision is demanded. Other off-screen sounds fall into the broad “ambience/immersion” category where, if anything, ambiguity of location is desirable. A sense of distance is an important factor in “transporting” listeners out of the listening room, and this factor is related to the essential perception of envelopment, the sense of being in a different, larger space.

8.2.2 The Subwoofer Channel: LFE and Bass Management

Many movies rely on massive bass sound effects for dramatic effect; it is one feature of movie sound, other than pure loudness, that most cinema attendees don’t have at home. Individual channels in the original film soundtracks are broadband, but the main cinema loudspeakers operate usefully only down to about 40 Hz. This was adequate until the first “blockbuster” films needed more low bass. Existing cinema systems were unable to handle the load without distortion and they lacked genuine low frequencies, so a new channel was created, called either “low-frequency enhancement” or “low-frequency effects” but always LFE. The bandwidth was restricted to frequencies below 120 Hz and the signal level was reduced

by 10 dB on the soundtrack to allow for more headroom, which was a problem with analog soundtracks of the day. The 10 dB is restored on playback. In cinemas dedicated subwoofers reproduce the LFE signal, which is used for extra impact in dramatic sound effects. This complicates the calibration process because the LFE content is not always correlated with the main channel signals and the bands overlap in the 40 to 120 Hz range. King and Theriault (2023) investigated this. This additional limited-bandwidth channel should be included in the "0.1" subwoofer channel in 5.1 or 7.1 and other domestic bass-managed configurations. Cinemas also should have a version of bass management.

Bass management is a signal processing option in multichannel receivers and surround processors with which it is possible to extract the lowest frequencies from those individual loudspeakers identified as "small", combine them, add them to the low-frequency effects (LFE) channel and deliver the combination signal to a subwoofer output and/or to full bandwidth loudspeakers (identified as "large") in the system. The normal crossover frequency at which this is done is 80 Hz. This can be changed, depending on system components. 80 Hz is popular because below 80 Hz it is difficult or impossible to localize the source if reflex ports don't "chuff" or if higher frequencies are present because the subwoofer is not steeply low-pass filtered, which may be the case with the standard bass-management filter. Some surround processors permit steeper low-pass slopes. Holman (1998) gives a good history. Figure 8.4 illustrates a common implementation, but there are other variations.

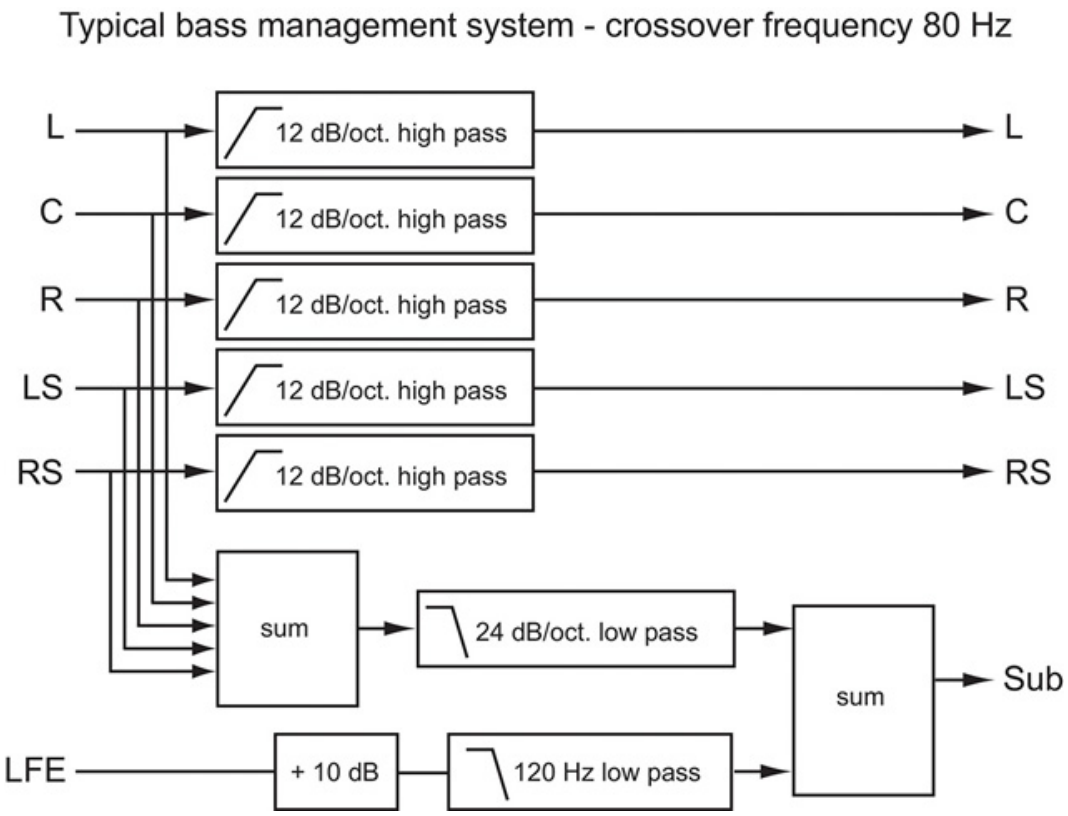


FIGURE 8.4 *A common version of bass management with a basic subwoofer-to-satellite crossover frequency of 80 Hz.* ⏎

Figure 8.4 illustrates one version of bass management in which the LFE (low-frequency effects channel in movies), containing frequencies up to 120 Hz, is allowed to pass unmodified to the subwoofer. Some implementations low-pass filter the LFE at 80 Hz before the subwoofer output, discarding some of the LFE signal. This is not encouraged. Neither are implementations that completely ignore the LFE. This information may or may not be available to consumers—one reason why different users may have different experiences with bass-managed systems.

A second reason for differing experiences lies in the frequency range between the subwoofer crossover frequency and the transition/Schroeder frequency around 200 to 400 Hz. This is the frequency range where *adjacent boundary* problems add to room resonances—topics discussed in later chapters.

Few can afford or have the space for the large floor-standing full-range loudspeakers that can reproduce the lowest organ pedal and synthesizer notes in music, and explosive impacts and thunder rumbles in movies. Employing bass management to separate the lowest frequencies in the program for separate reproduction through specialized subwoofers removes the need for massive front and surround loudspeakers. Relieving them of this task means that they can play louder with lower distortion.

Notions that subwoofers are somehow antithetical to good sound are mistaken. Smooth, extended bass is an important part of all entertainment, not just movies. However, the standard-issue bass management scheme does not have the flexibility to enable ideally smooth transitions in the crossover region in all installations. Some processors offer options to choose crossover frequencies, attenuation slopes, and perhaps delays and phase/polarity manipulations, all of which combined with in situ measurements are highly advantageous. Fortunately, steady-state room curves are reliable data at low frequencies. Equalization is likely to be required, and it works if used intelligently.

Todd Welti will explain this in detail later, including how multiple subwoofers can not only improve bass performance in notoriously problematic small rooms but deliver similarly good bass to multiple listeners. The solutions work in mono, stereo, multichannel and immersive systems and benefit any program with bass content.

8.3 HOW MANY CHANNELS, WHERE AND WHAT FOR?

We are told by name-brand companies where to place loudspeakers to reproduce their proprietary recorded content, and they are not all compatible with each other. Marketing claims aside, the differences don’t seem to matter to the entertainment value of the programs. The capacity of silver disc and digital database storage media allow for multiple soundtrack formats to be stored, but when the program is streamed, we are not always certain what format is being delivered, and how much channel upmixing or downmixing and compression (loudness and data reduction) has been

applied. Ignorance is sometimes a good thing, and programs continue to be enjoyed.

The following discussion begins with an attempt to summarize the scientific underpinnings of multichannel audio, relying on subjective and objective evaluations to provide guidance.

8.3.1 Immersive Audio Arrives

When digital technology allowed for enough channels and signal processing power to permit the audio soundstage to be expanded in height, several companies scrambled to come up with marketable solutions, all different in various ways. Dolby was close to the movie industry and delivered Atmos with great fanfare, debuting a 64-channel cinema system. This obviously offered film sound mixers exceptional flexibility in panning sounds to different loudspeakers in different locations, which was persuasive to most of the large audience. However, few cinemas can afford such an elaborate array of loudspeakers, and certainly homes cannot accommodate it. However, by using phantom images between and among pairs or groups of loudspeakers many more locations are possible with fewer loudspeakers.

Using “object-oriented” signal processing, engineers could pan sounds anywhere in the loudspeaker array, employing different combinations of loudspeakers to deliver phantom image impressions of localization at many locations between loudspeakers. A claimed advantage was that the system was scalable to installations using different numbers of channels/loudspeakers, reducing the installation costs for many small cinemas and, of course, home systems. But, as we all know from stereo phantom images, they are reliably localized for only a single listener in homes, or over a very small area of a large venue. Fewer loudspeakers present increasing challenges to maintaining correct perceptions of direction for more than one listener.

Naturally, the changing combinations of loudspeakers necessary to locate the “objects” generate changing patterns of acoustical interference. Consequently, sound quality varies with location. The author was a guest of Dolby, at their reference facility in Burbank, California, when Atmos was

debuted and experienced several demonstrations of cinema and home theater sound. The action movie excerpts were definitely dramatic and engaging. One demonstration in the cinema compared panning a male voice narration around the seven horizontal-plane loudspeakers in the conventional 7-channel mode and then in the Atmos mode. The conventional system behaved as expected with the voice jumping conspicuously from the front center loudspeaker to others as it moved around the room. In the Atmos demonstration the location changed much more smoothly, but the sound quality changed with it, being quite colored in some locations. When I mentioned this to the assembled senior management, there was a momentary pause while they looked at each other, and one then said something like "most people don't notice that". The next day some Harman colleagues went to the same demonstrations, but that panned voice example was inexplicably "not available". In fairness, this was a demanding test, a slowly panned male voice in isolation, but it was an excellent demonstration of an undesirable characteristic of object-oriented panning.

As discussed in Chapter 5, listeners are more accepting of sound quality flaws in a complex sound field. Perhaps when synthesized alien creatures or machines are flying overhead, considerations of sound quality are lost in the excitement of the moment in a movie. It is entertainment, but for music there may be disadvantages to object-oriented panning of positionally stable images. Movie mixers have learned how to employ fixed-position, loudspeaker-based images for important, stable content, and object-based images for transient, dynamic and spatial effects.

During my visit Dolby also demonstrated a comparison of four real ceiling loudspeakers and the upward-firing ceiling bounce version of Atmos for home theaters. The real loudspeakers were predictably much better. There were high-frequency height impressions from the ceiling-reflection version, but lower frequency sounds could still be localized at the loudspeakers. Directional array designs could be more persuasive, at higher cost.

DTS created a version of immersive audio, paralleling their competitive systems in the past, called DTS:X. Barco Auro3D was the product of Wilfried Van Baelen at Galaxy Studios in Belgium, claiming to be motivated by music, but also compatible with movie sound. The author heard Auro3D examples of classical symphonic and choral music recorded in concert hall and cathedral venues, and they were impressively realistic, even while moving around the home theater. There is also the European MPEG-H three-dimensional audio to be reckoned with, and it has some support as it is a proposed foundation of Next Generation Audio. All systems have incompatibilities in setting up a home theater, so compromise is often necessary.

8.4 MULTICHANNEL AUDIO SYSTEM OPTIONS

8.4.1 Channel Numbering Schemes

To keep track of how many loudspeakers there are in a multichannel system, and to help in understanding where they are located, the industry originally adopted a simple designation. It consisted of two numbers: the number of front channels and the number of surround/side/rear channels. So, 2/0 was stereo. 3/1 was the original Dolby Stereo/Surround system, with a single surround channel. 3/2 was conventional 5-channel surround, with L, C and R across the front, and two surround channels. When the existence of subwoofers was acknowledged, the numbering system was modified to 5.1, describing three front channels, two side surround channels, making 5 and a subwoofer channel, 0.1. This grew to 7.1 when two more surround channels, the rears, were added. Then came elevation channels so a number was added. Now, 7.1.4 describes a system with 7 channels arranged at approximately ear level, a single subwoofer channel (from which multiple subwoofers can be driven) and four upper-level loudspeakers or their ceiling-reflected approximations.

These days multiple subwoofers are increasingly being used to tame aggressive room resonances, delivering better bass to multiple listeners.

Most are connected to the same output, receiving the same signal but some advanced multi-subwoofer designs use processing in the line-level signals to individual subwoofers, so one might occasionally see 7.4.4 or the like. Elaborate immersive systems have even more channels, including front wide loudspeakers between the front L and R and the side surrounds, at ±50° to ±60°, and more upper-level channels. Different brands have different configurations.

In summary:

- The first number describes the number of main (ear-level) loudspeakers.
- The second number describes the number of subwoofer channels, but not the number of subwoofers connected.
- The third number describes the number of height loudspeakers.

8.4.2 The Basics of System Layout

There are several variations for laying out conventional listener-level multichannel systems, and even more with immersive systems. Figure 8.5 shows some popular floor plans. Small variations will not destroy the entertainment. It is the angles from the prime listener's perspective that matter, not the walls on which the loudspeakers are mounted; it depends on the shape of the room. For example, in (e) and (h) the rear loudspeakers could be located on either the side or rear walls of the room. The shaded areas are those within which small positional adjustments are possible without seriously compromising the results. In rectangular rooms, delays and sound level adjustments in the calibration/setup process compensate for differing distances to the reference seat.

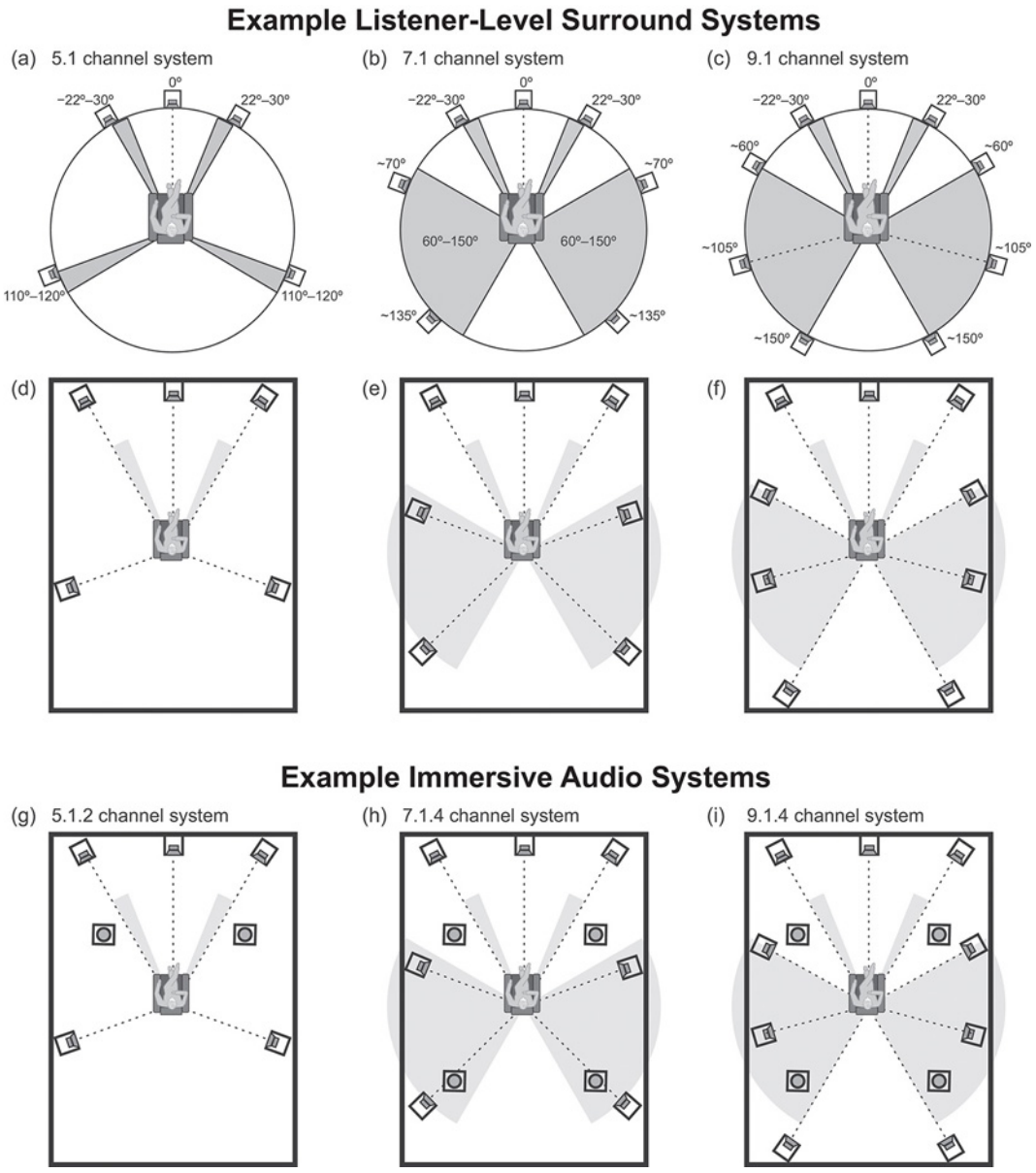


FIGURE 8.5 *Example loudspeaker layouts for various listener-level surround systems and immersive systems that add upper-level (in- or on-ceiling) loudspeakers. Only the satellite loudspeakers are shown. In the top row, (a), (b) and (c) illustrate the basic ITU-R BS.775–3 (2012) recommended arrangements for different numbers of channels, displayed in the "academic" circular setup. In the middle row, (d), (e) and (f) show how these arrangements can fit within four walls. In the bottom row, (g), (h)*

and (i) add some upper-level loudspeakers for immersive playback. ⏎

8.4.3 Dolby Atmos Music

Dolby is the major force in immersive audio, and their guidance to the professional audio engineers and artists who create the programs we listen to demand our attention. Dolby (2024) "Best Practices for Dolby Atmos Music Studios" states that it is "the industry expectation for commercial studios and reference environments and can be used as guidance for private mix rooms." In principle, if these recommendations are followed, one should hear very much what the creators of the art heard.

It is noteworthy that the loudspeaker layout includes front-wide channels at ±50°, which readers will recognize are effective locations for enhancing impressions of spaciousness and envelopment (see Section 7.1.2). They state: "9.1.4 or 9.1.6 configurations, which utilize 'wide' speakers are preferred and considered our standard reference". This is a good idea.

The recommended loudspeaker layout dimensions can easily be accommodated within typical domestic listening rooms and home theaters. The sound level expectation is a peak of 105 dB, but it seems to be for the entire system, not individual channels as in cinemas, and reduced levels are permitted for smaller installations, or more casual use. LFE is included, and bass management permitted where subwoofers are necessary to supplement small loudspeakers. Bass extension is to 31.5 Hz and the crossover frequency is 100 Hz, which is a curious choice, because of the risk of localizing the subwoofers. However, that risk is minimized by the assumption that they will be located on the front wall with the L, C and R. They don't plan on any front-back multi-subwoofer room-mode manipulation, relying on equalization averaged around the mix position. A missed opportunity.

There is a focus on the "translation" of mixes among recording facilities. This is important, but achieving it requires uniformity in performance, not

excellence in performance. Performances must also “translate” to consumers.

The requirements for the loudspeakers are insufficient. They specify a generous ±3 dB tolerance on the on-axis frequency response as specified by the manufacturer, not always a reliable source, and few publish meaningful curves. It matters little because a single curve is inadequate to describe loudspeaker performance. Section 2.9.1 describes one example of how misleading such a specification can be. This criterion allows a wide range of loudspeakers to qualify, including poor performers. Current loudspeaker practice is to use a comprehensive set of anechoic curves, or the spinorama —an industry standard—to describe loudspeaker performance. From these data reliable estimates of sound quality and steady state room curves are possible above the transition frequency. Below it, small room resonances are dominant. Measured room curves are results, not targets, because equalizing a flawed loudspeaker to have the same room curve as a good one is no assurance of good sound quality.

Dolby proposes the target room curve shown in Figure 8.6 as a standard for the production of Atmos music for the world. The fact that it approximates the anticipated room curve from highly rated loudspeakers raises hope, but there is a problem. Steady-state room curves are primarily determined by sound radiated far off-axis, which is not included in the on-axis selection criterion for the loudspeakers. Loudspeakers vary considerably in their directional behavior. Consequently, an unknown characteristic of the loudspeaker is equalized to match a target room curve, changing its very important on-axis, direct sound, response in the process. This was the criterion for choosing the loudspeaker. Adding further confusion, there is allowance to alter the high-frequency tilt if it doesn’t sound right for the room; this is a tone control, not a standard. If it is expected that this will result in high and uniform sound quality throughout Atmos music facilities and “translation” to listeners outside, it cannot work.

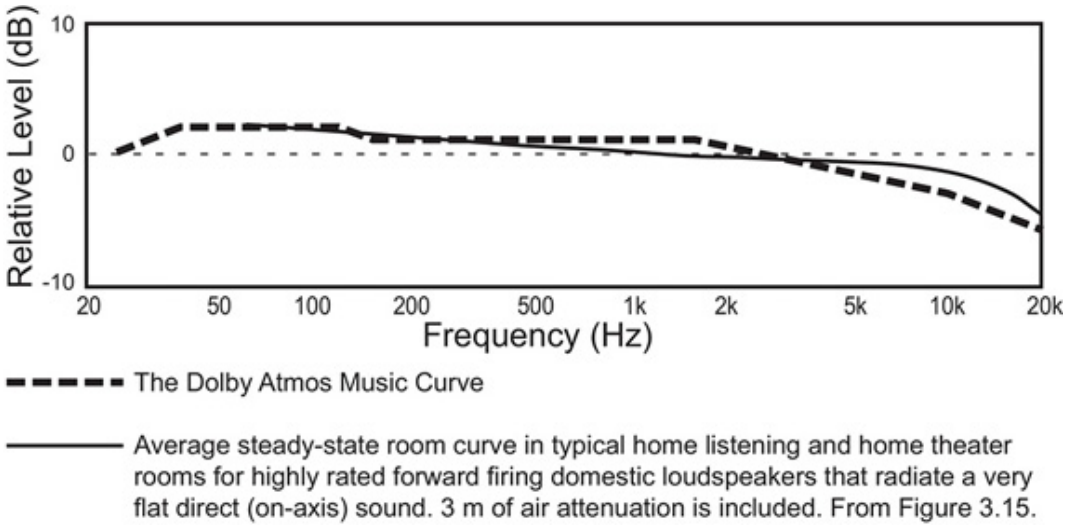


FIGURE 8.6 *shows the steady-state Dolby Atmos Music Target Curve compared to the average steady-state room curve measured in small rooms from highly rated domestic forward-firing loudspeakers—from Figure 3.15.* ⏎

This is a disappointing set of recommendations from a company with so much influence in the audio industry. It is unfortunate that Dolby didn't use the opportunity to set a modern, meaningful and high standard of performance for sound quality. They say: "The curve was developed in partnership with studios and music labels who required confidence in translation across multiple monitoring and playback environments." What has been suggested follows industry traditions, not current science.

8.5 SELECTING THE APPROPRIATE LOUDSPEAKERS

It should now be evident that the starting point for any sound reproducing system is well-designed, timbrally neutral, loudspeakers. Readers of earlier chapters will by now be able to recognize such loudspeakers by inspecting spinorama or comparable comprehensive anechoic data. Fortunately, such loudspeakers are now available from several manufacturers, from several countries, in many sizes, configurations, and styles. It may take a little internet searching to find them, but it is worth the effort. Section 2.9

contains links to some sources of trustworthy spinoramas and other anechoic measurements.

If conventional forward-firing loudspeakers are used it is recommended that they are all aimed at the prime listening location to deliver a pristine direct sound, the anechoic data will show how much tolerance there is. This ensures that there will be one very happy listener, and probably others seated nearby. Mounting loudspeakers on swivel brackets in surround locations can be challenging. They are not especially attractive, and by suspending the loudspeaker close to a wall there is an adjacent-boundary acoustical interference (adjacent-boundary problems will be covered in a later chapter). Custom home theaters attempt to find clever ways of hiding them on or in the room boundaries.

Popular alternatives to these are on- or in-wall or ceiling loudspeakers. Some of them have very wide dispersion, needing no aiming, and others offer flexibility in the aiming of the tweeter. As mentioned in Section 7.3 the elevation cue for localization is in the high frequencies above about 6 kHz so one would want ceiling-mounted loudspeakers with tweeters that can deliver these frequencies to as many seats as possible. Anechoic off-axis ("early reflections" in a spinorama) and/or directivity index information is important. In- and on-wall or ceiling loudspeakers are visually much less obtrusive, and if care is taken in choosing and installing them, they can deliver excellent sound quality. The best ones are comparable in sound quality to free-standing loudspeakers, and they avoid one adjacent-boundary problem by being mounted in it. Some have back-boxes, providing a known enclosure for the woofer, but others are designed to work in well-damped—fiberglass-filled—wall cavities. Used with bass management and a subwoofer system, as discussed in Section 8.2.2, they are not required to reproduce low bass.

8.5.1 Loudspeaker Directivity Requirements

Sound from loudspeakers is of little value if it is not delivered to listeners. In stereo, one aims the loudspeakers at a listener and the job is done. In home theaters it is more difficult, depending on the room layout. Figure 8.7

shows a difficult situation, a crowded room. Ideally, one would not want any listeners close to walls, and close to surround loudspeakers that can be distractingly localized when the sounds they radiate are intended merely to add ambience, audience sounds, etc. The larger the room and the audience, the greater number of loudspeakers necessary to deliver a satisfying experience to most of the listeners. There will always be some seats for those guests who fall asleep even during a good movie.

The first task is to ensure that all listeners are delivered a strong direct sound from all loudspeakers. This localizes the sound event, which is especially important in immersive systems, and provides the dominant cue for perceived sound quality. Early reflections are the second loudest sounds listeners hear in small rooms, and they cannot be ignored. Their effects have been discussed in earlier chapters, and there are both positive and negative attributes, depending on the off-axis performance of the loudspeakers, the absorptive or diffusive nature of the reflecting surface, the program material and, of course, personal preferences. The treatment of first-reflection surfaces, especially side walls, is optional and it requires some thought. The slavish attraction to automatically absorbing all reflections is something that had its origins in listening to stereo in recording control rooms, where the requirements of the recording engineer to hear tiny details of his/her manipulations surpasses all other objectives. At home, they, and most people, prefer some reflected sound for recreational listening. With multichannel audio the “large room” recorded room sound dominates, as was discussed in earlier chapters, especially Chapter 5.

Figure 8.7 contains a few useful messages. This is a large, dedicated home theater, with many people to entertain. Smaller installations will be less difficult, but all need some thought. The front loudspeakers seem not to present any challenges from the point of view of dispersion for audience coverage if the L and R loudspeakers are aimed at the prime listening location. Conventional forward-firing designs will work well. However, if the surround loudspeakers are forward-firing designs, they need to be aimed in order to cover the audience. If they are in- or on-wall designs, they must have very wide dispersion. The dispersion in loudspeakers is symmetrical,

whether needed or not, in which case some sound energy will be radiated in unproductive directions, adding to the reflected sound field, which is mostly unwanted. Knowing this, additional acoustical absorption may be advisable on the surfaces receiving the excess sounds. Scale drawings like the one shown are very helpful.

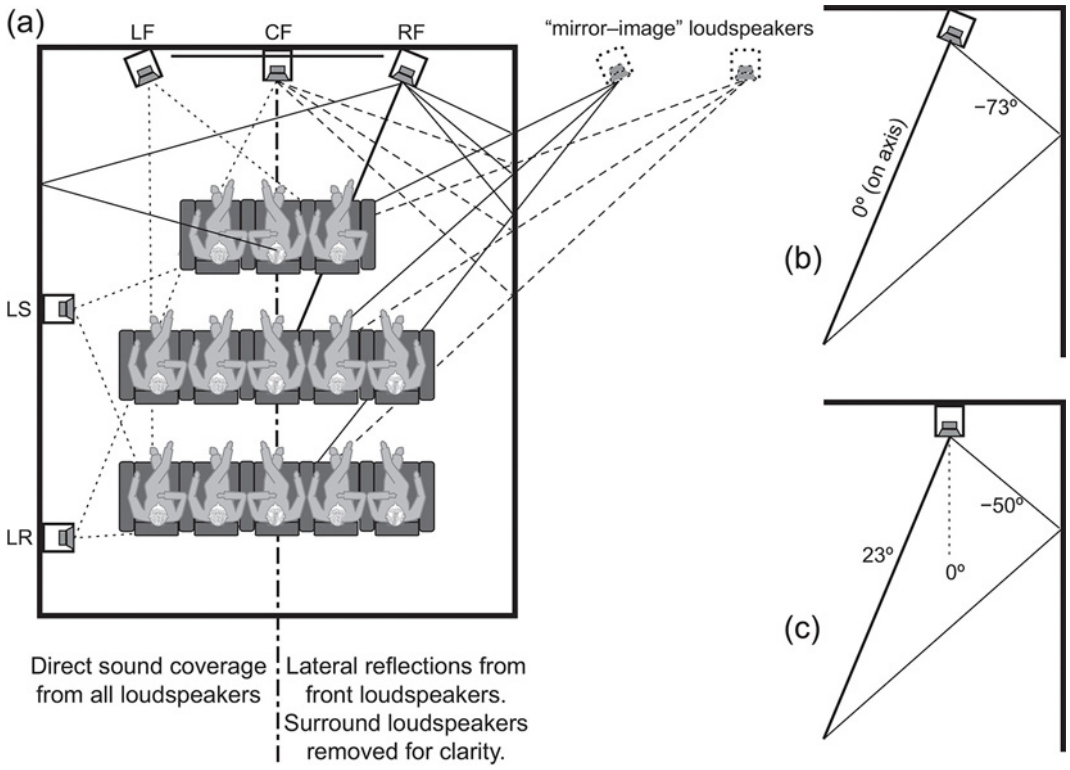


FIGURE 8.7 *(a) A home theater floor plan showing on the left half the angular spread from the loudspeakers necessary to deliver direct sounds to all listeners. The right half shows the first laterally reflected sounds delivered by the center and right channels to center seats in each of the rows. (b) and (c) show the effects of two loudspeaker aiming options, one at the listener (b) and one straight into the room (b).*

The difference in distance between front and rear rows will result in sound level differences, which will result in different experiences. From the front row to the back row in this example, the direct sound will fall according to the inverse-square law (−6 dB/double distance, explained in a

later chapter) being reduced by about 8.4 dB, approaching half loudness. However, the reflected sounds in normal domestic spaces add to the direct sound, and the rate of decline is reduced to about −3 dB/double distance, so the steady-state sound level is reduced by 4.2 dB. Quality entertainment is possible in all rows, but the experience will be different because moving away from the front loudspeakers diminishes the proportional contribution of the important direct sound to the total listening experience. Systems are calibrated for the prime listening location, in cinemas it is a large area centered 2/3 of the way back from the screen in the center portion of the room, so get there early.

The off-axis angles at which side-wall reflections occur should be noted. The right half of Figure 8.7(a) shows the sound rays for sounds arriving at the center seat in each of the rows. For the front row, the off-axis angle is about 89°, an extreme angle at which much reduced mid- and high-frequency energy is radiated. Data for conventional forward-firing loudspeakers indicates typical reduction of about 10 dB from about 700 Hz to about 4 kHz, increasing rapidly above that. To this reduction is added the inverse-square propagation loss, and high-frequency air attenuation. For the front row the side wall reflection is not an issue. For the middle row, probably the prime listening location, the off-axis angle is about 73° where the radiated mid- and high-frequency energy is reduced by roughly 6–8 dB in mid frequencies, and more above. For the rear row the angle is about 63°. If absorption is chosen, 2- to 4-inch (50 to 100 mm) fiberglass or acoustic foam slab panels are effective above the transition frequency. Obviously, loudspeakers with different polar radiation characteristics will need to be treated individually, see Figure 4.5.

Figure 8.7(b) and (c) contrast what happens with different loudspeaker aiming. In (b) the listener benefits from the best on-axis direct sound that most well-designed loudspeakers radiate, as well as much-reduced reflected energy from the side wall. Because a loudspeaker is in a rectangular enclosure with a flat back does not mean that it should be oriented with the back parallel to the wall—visual appearances notwithstanding. In-wall front loudspeakers need to be installed in angled baffles. This alignment affects

the L and R loudspeakers, which are important if listening to stereo is to be an option.

In setting up a listening room it helps to start with some squared paper and make a scale drawing of where loudspeakers and listeners are located. A computer-based approach is an alternative but may take longer than a pencil and eraser. These are rough estimates, and great accuracy is not required. It will also be useful when deciding on seating and subwoofer configurations for improved bass in small rooms—an essential operation discussed in detail in a later chapter.

Figure 8.8 shows that any competently designed loudspeaker easily meets the task of the front channels when the left and right loudspeakers are angled inward, which should be normal practice. The surround channels are more challenging. If the loudspeakers are on adjustable brackets, they can be aimed to maximize coverage, but such mounting schemes are unsightly and place the loudspeakers at a distance from the walls causing a boundary interaction. In-wall or on-wall designs are much better as they avoid the acoustical interference problem with the adjacent boundary.

The challenge is to find loudspeakers with the necessary horizontal dispersion to cover the audience. In general, this is satisfied by bipole—bidirectional in-phase—designs. There are other variants with clever names, but most amount to no more than loudspeakers having very wide dispersion—a good thing in this case. Dipoles should be avoided.

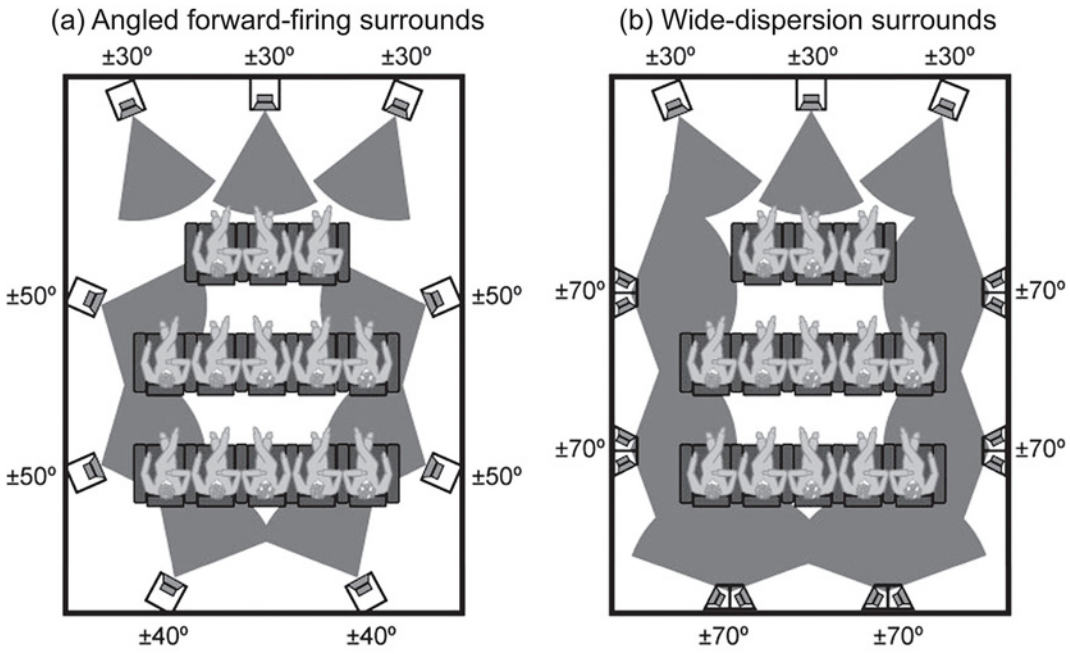


FIGURE 8.8 *A summary of the horizontal-plane angular dispersions required of the loudspeakers to deliver direct sounds of comparable quality and level to all. The criterion of excellence for direct sounds that they should all be as similar as possible to the on-axis performance of the loudspeaker. This is obviously a challenge for the surround loudspeakers if they cannot be aimed, as shown in (a). In (b) the surround loudspeakers are all non-steerable in-wall or on-wall designs, and it can be seen that much of the sound is radiated in unproductive directions. A combination of the designs may be optimal, depending on circumstances.* ⏎

Height and ceiling loudspeakers in immersive systems need similar directivity considerations as it is important for frequencies above about 5 kHz to be reliably delivered to listeners as these provide the localization cues for elevation, see Figure 7.10.

8.5.2 Selecting Front, Left, Center and Right Loudspeakers

Longstanding stereo tradition has instilled a belief that any audio system of value must start with a pair of full-range floor-standing loudspeakers. However, if one is aiming to get a system that can provide both stereo and

multichannel reproduction, and that can entertain multiple listeners with movies, music concert videos, as well as audio recordings, then a different approach is required. A well-designed multichannel home theater can also be a well-designed stereo system; no compromise is necessary.

The only restriction is that forward-firing designs are easier to incorporate into such systems. For those who find unique virtue in other directional designs, there is a problem. These other designs usually offer different radiation patterns aiming to embellish the stereo listening experience, to add spaciousness. The only consolation I can offer is that the additional channels in a multichannel system more than compensate. Multichannel music recordings can be remarkably gratifying, and the right upmixer, adjusted to one's liking, can embellish much 2-channel source material in ways worth experiencing. The fallback is always stereo in its natural form.

All multichannel receivers and processors incorporate bass management, so enormous full-range loudspeakers are not required. One can still have floor-standing loudspeakers, but they need not be the top-of-the-line behemoths that are capable of subwoofer frequencies. Let the subwoofers do what they do best. The reality is that, because of small-room acoustics, the necessary locations for L, C, R loudspeakers are not where we want to position one or more subwoofers to entertain multiple listeners with non-resonant, clean, tight bass. This is an important issue that Todd Welti will be discussing in detail in Chapter 14.

For now, let us focus on loudspeakers that function well above about 80 Hz, the most common subwoofer-to-satellite crossover frequency. If floor-standing loudspeakers are chosen for L and R channels, as is often the case, they are likely to be able to deliver frequencies lower than 80 Hz, in which case when they are high-pass filtered by bass management they will be able to play louder with lower distortion. Just don't waste money on excessively large units.

The earlier part of the book has adequately described the measured performance goals of conventional, vertically arranged transducers in floor-standing or satellite/bookshelf loudspeakers. Flat and smooth on-axis and

listening window frequency responses are necessary starting points. Smoothly declining early reflections and sound power curves are additional positive cues. These properties ensure smooth, gradually changing directivity indexes. A little practice and guidance from the online sources that measure spinoramas will help. It is important not to forget that bass accounts for about 30% of overall sound quality ratings, and that is dominated by the listening room. For matters of adjusting spectral balance in programs to taste, do not hesitate to use tone controls: that is what they are for.

Historically, the center channel loudspeaker has been treated overly casually, often a low-cost small box to fill a need. In movies, the center channel does most of the work, providing most of the dialog and much of the on-screen action sound. In TV it is well used, depending on the service provider and channel. Talking heads should be delivered through the center channel, but, sadly, this is not always done. Stereo traditions persist.

In cinemas, all three front channels are identical loudspeakers. Acoustically transparent projection screens in home theaters allow for physically identical L,C,R loudspeakers, but with direct view displays the center loudspeaker may need to be reconfigured to have a lower vertical profile and positioned below or above the screen—usually below in homes. In home systems we need to focus on finding a center loudspeaker that matches the acoustical performance and sound quality of the L and R units. Many manufacturers have center channels as part of a compatible set, making the choice an easy one. If this is not the case, then one truly needs to have anechoic data on all the loudspeakers in order to ensure that there will be a close match among the three front units.

Modifying transducers into a horizontal configuration adds a complication. Conventional loudspeakers have woofers, midranges and tweeters arranged vertically, thereby ensuring a uniformly wide horizontal dispersion of the sound. There will be irregularities in vertical dispersion because of the spacing of the transducers, but this is not a major perceptual issue in normal rooms. Center-channel designs can be more challenging.

In Figure 8.9(a) it is seen that the MTM (Midrange-Tweeter-Midrange) design has two “midranges”, which actually are woofer/midranges. These acoustically interfere with each other at increasing horizontal angles because they are physically separated and both radiate sound up to high frequencies in order to cross over to the tweeter. This can be seen in the polar plot displaying a horizontal view of sound level as a function of frequency and angle. Moving radially from the center the concentric rings represent 200 Hz, 500 Hz, 1 kHz, 2 kHz, 5 kHz, 10 kHz and 20 kHz. The areas in white represent sound levels within 3 dB of the axial (0°) output. The contours at the onsets of progressively darker gray areas represent sound levels at 6, 12, 24 and 48 dB below the axial output. Moving horizontally off axis the first interference dip in the frequency response appears at under 10° off axis. Recall that in Figures 8.7 and 8.8 it was shown that a horizontal dispersion of ±30° was required for the center loudspeaker to deliver intact direct sound to all listeners in a typical home theater. This figure shows that by 30° this loudspeaker is experiencing heavy acoustical interference and the output has dropped seriously over a wide frequency range. This is not good.

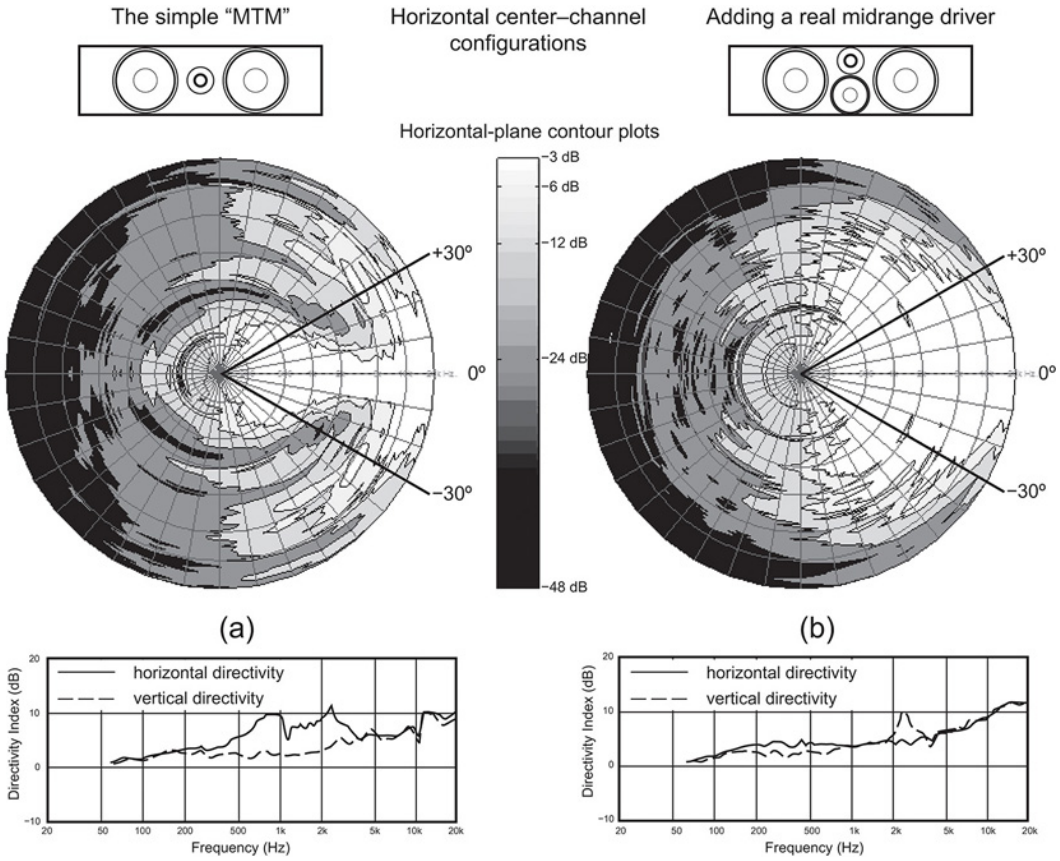


FIGURE 8.9 *Two common configurations for horizontal center-channel loudspeakers showing horizontal sound-level distribution displays and horizontal and vertical directivity measurements.*

Figure 8.9(b) depicts a much better design, in which a midrange loudspeaker has been added in a central location. This allows the widely spaced woofers to be turned off at a lower frequency, and the interference dip disappears. The superior dispersion of the small midrange driver is apparent. The result is that at 30° the sound is unimpaired; only a normal slight loss above 10 kHz due to tweeter directivity is seen (this would be seen in any loudspeaker with a conventional dome tweeter). The bottom graphs show separate vertical and horizontal directivity indexes for these designs. The horizontal dispersion problem with the MTM layout and the virtue of the three-way design are apparent.

8.5.3 Selecting Surround Loudspeakers

These channels provide impressions of both direction and space. In movies sounds are momentarily localized to off-screen locations. The locations are non-critical; general left-right, front-back directions are sufficient, because only that is possible in traditional cinemas with multiple surround loudspeakers. Immersive soundtracks are more capable. In music they may be called upon to provide stable images of vocalists, musical instruments, etc. and again, great precision is not a requirement. Most of the time, these loudspeakers provide ambient sounds, atmospheric mood-setting music and other sounds that are intended to float ambiguously in space. This illusion is associated with the impressions of spaciousness or envelopment like those experienced in large halls.

The notion that envelopment is perceived in a highly diffused sound field like a concert hall has led to a belief that a diffuse sound field is a requirement. It isn't. As shown in Section 7.2 a credible perceptual facsimile can be created using a small number of channels and loudspeakers. Impressions of spaciousness and envelopment exist when the sounds arriving at the listeners' ears have random differences in amplitude and time, i.e. they are uncorrelated, a condition that is maximized by sounds arriving from the sides, and not by sounds arriving from angles close to the front-back axis.

In the present example shown in Figure 8.8, for delivery of the direct sounds the surround loudspeakers require uniform horizontal dispersion out to ±70°, beyond the capability of conventional forward-firing designs. A simple technical description of such a loudspeaker would be a horizontally omnidirectional (*very*-wide-dispersion) on-wall or in-wall loudspeaker.

An approximation of this kind of loudspeaker is shown in Figure 8.10.

Figure 8.10(a) shows a spinorama measurement for one side of the bidirectional loudspeaker. It is very well behaved. (b) shows the in-phase, bipole setting and it also is well behaved, with greater sound output because both sides are operating in unison. (c) shows the out-of-phase, "dipole" setting result. To preserve low-frequency output the system progressively reverts to in-phase operation below about 500 Hz.

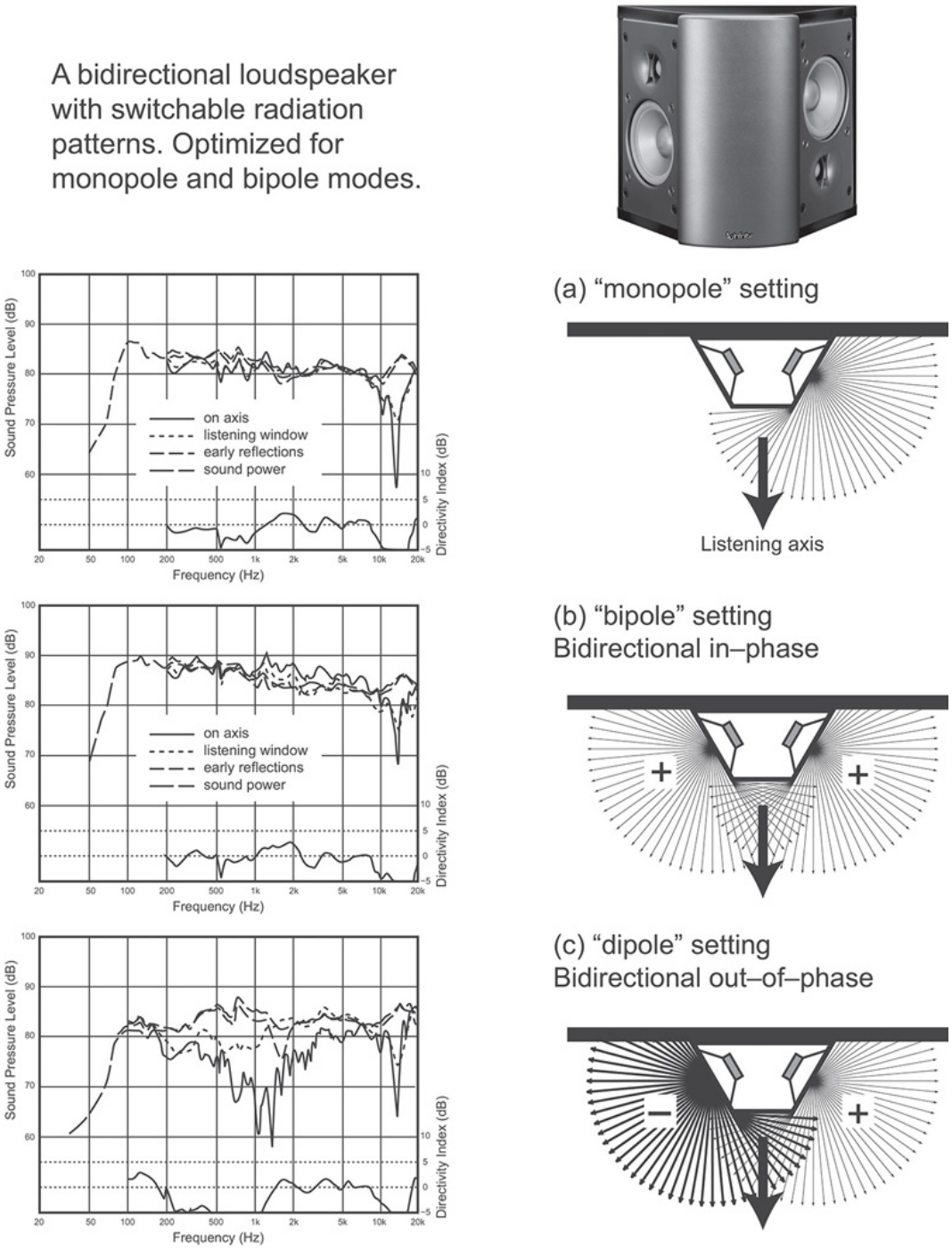


FIGURE 8.10 *The Infinity Beta ES250 (obsolete), a bidirectional on-wall loudspeaker measured in anechoic half-space, in each of its switchable directivity modes. The listening axis is perpendicular to the wall.* ⏎

In the monopole and bipole modes this is a good sounding loudspeaker, and in the bipole mode it has the acoustical output capability and wide dispersion necessary to meet the surround loudspeaker needs described in Figure 8.8. It is an example of good purpose-driven engineering.

In contrast, the measured performance in the dipole mode is unimpressive. Why is it there? Because at the time there was a market demand for "dipole" surrounds, generated by THX, and marketing thought it unwise to have a product that did not offer the option. But all were delivered to customers set to "bipole" to encourage use in that mode.

The dipole mode needs explanation. Back in Section 8.1.1, when discussing the evolution of multichannel audio, there was a period in which we lived with four matrixed channels, L,C,R and one channel for surround, Dolby Surround and Dolby ProLogic. In cinemas with many loudspeakers around the sides and rear of the large auditorium there was a pleasing effect. But when reproduced through two side loudspeakers in a small room it was not as rewarding. One of attempts to alleviate the problem was the "dipole" surround, a bidirectional out-of-phase (polarity-inverted) spaced pair of transducers. A true dipole has no separation of the opposite polarity radiating surfaces, as in front and back of a diaphragm or thin film radiating surface. These were not true dipoles. The goal was to attenuate the direct sound from the side loudspeakers and send more sound into the reflected, "diffuse", sound field in the room. It sounds reasonable, and THX promoted it energetically. It sounded "different" and it acquired a following, traces of which remain today—remarkable in the era of Atmos with numerous discrete channels not requiring any "assistance". The following is a technical analysis for the record.

The loudspeaker shown in Figure 8.10 was not optimized for the dipole mode as conceived by THX. The direct sound represented by the on-axis frequency response exhibits a large irregular mid-frequency depression instead of a broadband timbrally neutral "null" as might have been expected from a true dipole.

For completeness, let us see examples of four dipole surround loudspeakers that were designed for that sole purpose, three of which were

approved by THX and for which they received a licensing fee, Figure 8.11. Little needs to be said if one attends to the bold solid black curves that describe one of many possible direct sounds from these loudspeakers; they are all unattractive and very different from each other. One wonders what the criterion for approval was. Sound quality was obviously not a consideration.

The large difference between the direct sound and the listening window curves indicates that the null is very narrow, meaning so it is effective only for a single row of listeners with their heads in a line 90° perpendicular to the side wall. These loudspeakers must be located at ±90° to that row of listeners. Much of the sound energy that is radiated reflects off the front and rear walls of the room, which are poor locations for enhancing spatial effects, see Figure 7.4(b). Small acoustically damped domestic rooms cannot support a diffuse sound field of any magnitude so that is not a factor.

In terms of specifying the performance of these loudspeakers it was common practice to use sound power, the total sound radiated by the loudspeaker. This can only have validity in a strongly diffuse sound field, which is impossible in small listening rooms. Of these, only (c) appears to have achieved a ragged approximation of that target performance. In typical listening rooms the sound quality from these loudspeakers is predictably colored.

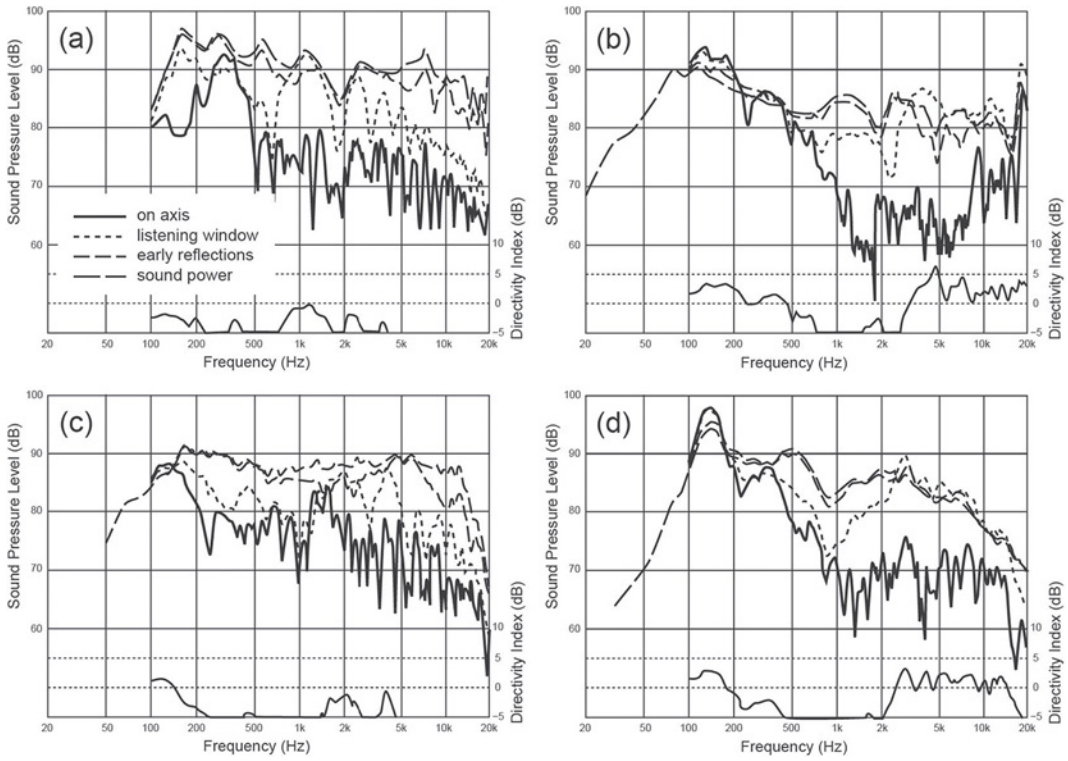


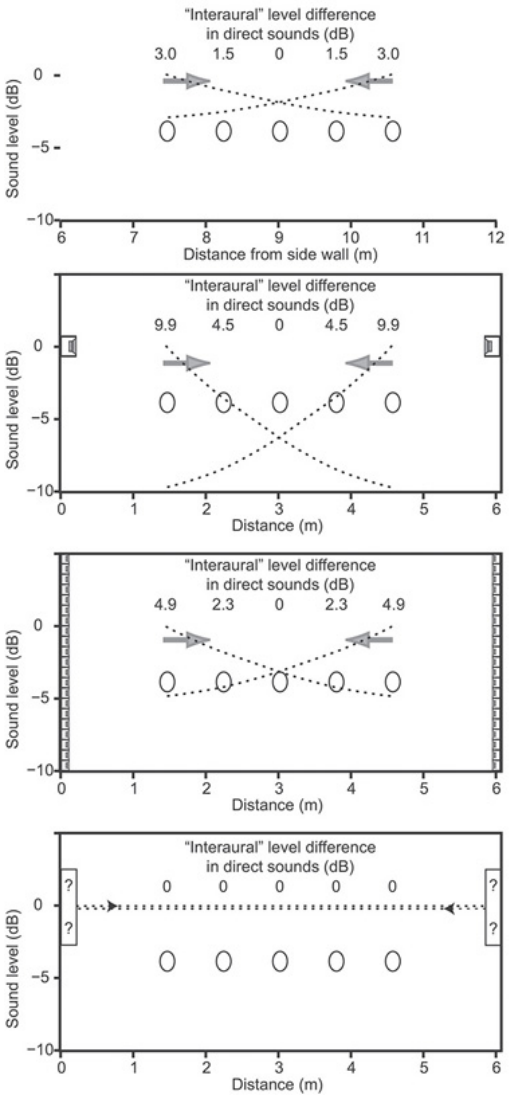
FIGURE 8.11 *Four "dipole" surround loudspeakers designed for the purpose.*

All considered it is evident that this misadventure was not the result of applied science. Dolby ProLogic has been replaced by digital discrete systems. Two of the strong proponents of dipole surrounds back in the day still manufacture loudspeakers and there are no dipole surrounds in their offerings, just forward-firing designs and bipoles. End of story.

8.5.4 A "Physics" Problem With Side Surrounds

It has by now been well established in this book that envelopment is an important perception, and that it is maximum when the sounds arriving at the two ears are of similar level and uncorrelated—different—i.e. low interaural cross correlation (IACC). The important differences are those that are in recordings, originating as reflections in the recording venue, captured by microphones or electronically synthesized to simulate reflections in a real space.

We know that the recording engineer seated in a prime location heard these sounds in their correct proportions. We know that the listener at home in the prime location is likely to hear them, as well as other listeners on the front-back center line of the room. But what about listeners seated to the sides of center line? Ideally these sounds should be properly delivered to all listeners, but they are not.



Surround loudspeaker propagation loss issues: good envelopment requires similar sound levels delivered from both sides—to all head locations.

(a) In a 60 ft (18 m) wide cinema equipped with the normal "point source" surround loudspeakers: conventional forward–firing cone/dome or cone/horn systems, located beyond the limits of the diagram. We examine five center–of–row seats.

Direct sound: −6dB/double distance

(b) In a 20 ft (6.1 m) wide home theater equipped with "point source" loudspeakers:
e.g. all conventional small–box systems
with monopole, bipole, or dipole horizontal patterns.

Direct sound: −6dB/double distance

(c) "LIne source" loudspeakers: must be very close to full floor-to-ceiling length.

Direct sound: −3 dB/double distance

(d) The "perfect" surround loudspeaker: constant level from both sides at all listening locations.

Direct sound: no (apparent) propagation loss

FIGURE 8.12 *(a) The anticipated sound levels arriving at five listening locations halfway across the width of a cinema, from conventional surround loudspeakers to the left and right. (b) The comparable situation in a home theater. (c) The effect of changing*

the surround loudspeakers to floor-to-ceiling line sources. (d) What we really want. ⏎

Sounds are maximally different, IACC is lowest, when they arrive from the sides, but there are problems delivering sounds of similar level. Figure 8.12 illustrates that persons seated midway across the room will always perceive good envelopment if the information is in the recording. However, at seats away from center the interaural amplitude differences increase, envelopment is degraded and as the side wall is approached, the side loudspeaker can be localized as a source and the effect is completely lost. There truly is a sweet spot in the room.

The reason is in the physics of sound propagation. The well-known inverse-square law tells us that the sound from a source that is small compared to the wavelength being propagated is reduced in amplitude by a factor of 6 dB for each doubling of distance from the source (−6 dB/dd where "dd" is double distance). This is true for all sounds in reflection-free (anechoic) environments and for the important first-arriving direct sound in rooms. This relationship changes to −3 dB/dd for floor-to-ceiling line source loudspeakers—not "tall" loudspeakers that are at best truncated lines sources, very different devices.

Figure 8.12(a) shows us that in a large cinema, five listeners seated around the center between the walls experience relatively small interaural differences in direct sounds from the two sides: 0 dB at center and 3 dB at the side seats. Probably they will all be pleased.

Figure 8.12(b) shows that in a much smaller home theater the consequence of propagation loss is dramatically greater: still 0 dB at center, but 9.9 dB at the side seats. Those listeners will be localizing the nearby side loudspeakers as sound sources, and envelopment is sacrificed.

Figure 8.12(c) shows what happens when conventional loudspeakers are replaced by vertical line-source loudspeakers running from approximately floor to ceiling. The propagation loss is reduced by half and so is the interaural difference in direct sound: 0 dB at center and 4.9 dB at the side seats. The side listeners may or may not localize the loudspeaker as a

source, but anyone who has experienced this kind of full height line source knows that they do not draw attention to themselves.

Figure 8.12(d) shows what we would like to have. No interaural differences in direct sound at all seats. Impossible? Perhaps not.

Professional audio engineers are constantly seeking loudspeaker designs that provide controlled coverage of large audiences. Traditionally, this has been done using large horns with varying directivities, line sources and now computer designed, digitally controlled line arrays. One of the clever innovations, the constant beamwidth transducer (CBT) was described by Keele and Button (2005). It is capable of delivering remarkably uniform sound levels over large areas and large distances. I have experienced walking from the far end of a large room up to the loudspeaker, ending up standing close to it, and hearing very little sound level variation along the way. It seemed to defy physics, but in fact it is a clever application of physics. The design will be discussed again in a later chapter, dealing with adjacent boundary effects, but for now, I have taken some liberties with Keele and Button's data and created Figure 8.13.

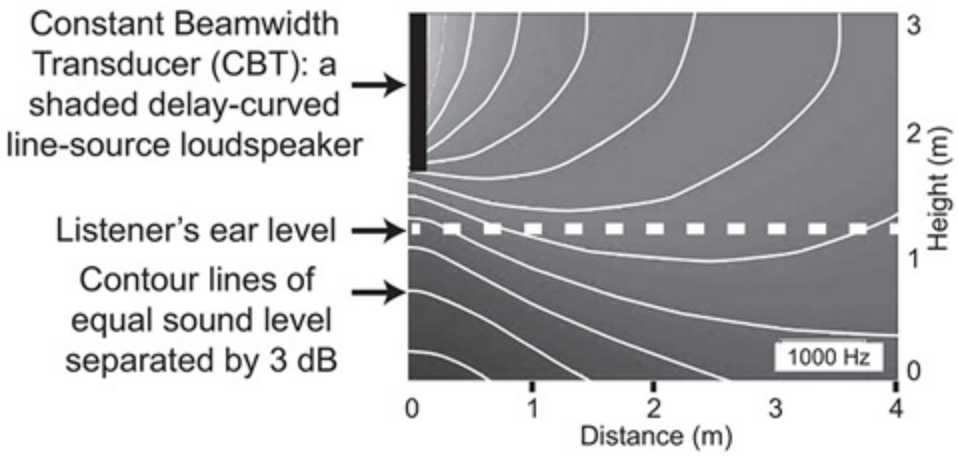


FIGURE 8.13 *Data from Keele and Button (2005) reconfigured by the author. It illustrates the result of inverting the mounting of the CBT loudspeaker, placing it adjacent to the ceiling, instead of the floor, because it results in more uniform sound level vs. distance at the ear level of listeners across the width of the room.*

Figure 8.13 shows that at 1 kHz a CBT loudspeaker designed to work with a reflective boundary, like a floor or ceiling, is capable of delivering sound levels to listeners within a 3 dB tolerance over a 4 m distance in this simulation at 1 kHz. Different frequencies will have different results but the overall effect is relatively consistent sound level over large distances. It approaches the "ideal" curve in Figure 8.12(d). Could this design be the optimum surround loudspeaker configuration?

JBL Professional has taken the CBT concept further, developing a line of versatile CBT audio products for general use as sound reinforcement and playback applications. Logically they have been used in custom home theaters and cinemas as surround loudspeakers. There is a downloadable calculator allowing designers to preview the audience coverage before installation (https://jblpro.com/resource/constant-beamwidth-technology-cbt-series-passive-controlled-coverage-columns.pdf).

To summarize, it is well established that the perception of spaciousness and envelopment is a major motivation for multichannel audio. Figure 8.12 shows persuasively that this is difficult to deliver to multiple listeners in small rooms unless one employs loudspeakers with less than the common inverse-square-law propagation loss. Long line sources are an improvement, and CBT loudspeakers are an interesting option.

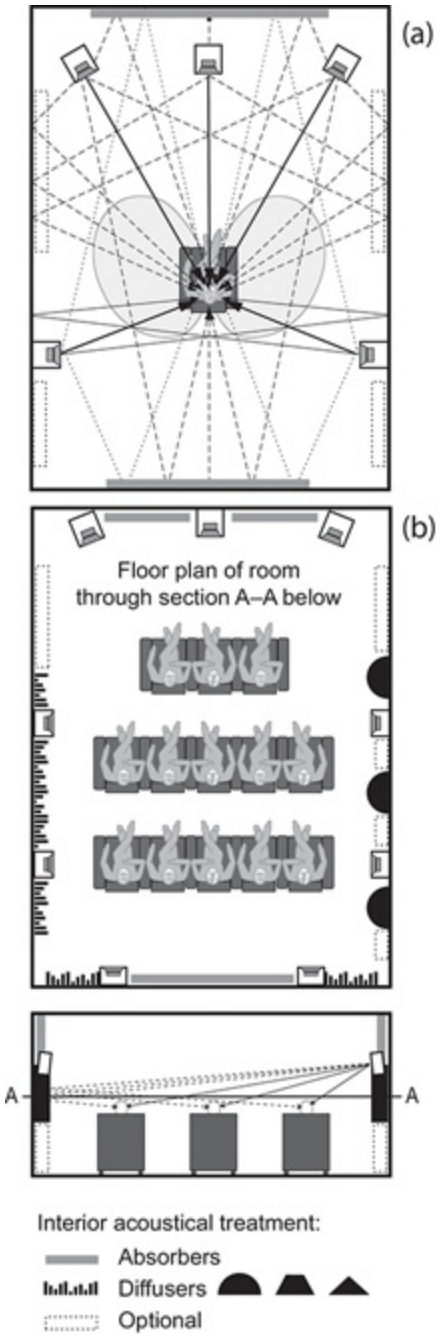


FIGURE 8.14 *Shows possible acoustical treatment arrangements that seem to contribute positively to listening experiences. (a) is for a small room small audience system; (b) is for a large room large audience system.* ⏎

8.5.5 Acoustical Treatments Can Help or Hinder

Figure 7.4(b) showed that direct sounds from the side surrounds and reflections of those sounds from the opposite wall both fall comfortably into the "balloon" and therefore contribute to impressions of spaciousness. If the direct sounds are diminished by propagation loss, it can be of some assistance if strong reflections of those sounds are possible from the opposite wall. It is not a perfect solution, but it costs nothing. The worst thing to do, it seems, would be to spend money on acoustical absorbers for those locations, aggravating the situation.

Figure 8.14(a) illustrates a suggested plan for acoustical treatment in a small room, based on Figure 7.4(b), showing the "spaciousness balloon". It shows bare walls to reflect the side surround loudspeaker sound, assisting the problem described in Figure 8.12. Because acoustical absorption is needed to achieve a desirable reverberation/reflection decay time in the room, the suggested locations are those where reflections contribute least to the spatial impressions, on the front and rear walls. The front side wall treatment is shown as optional, following earlier discussions that indicate that while there might be some negative impressions from them, if the loudspeakers are well designed, having relatively uniform off-axis (early reflection) performance, these reflections are found to be desirable to many listeners. Thick (3-inch) absorbers will essentially eliminate them, thinner ones change the sound quality, engineered or curved surface diffusers will attenuate and disperse them, and a bare wall reveals them. It is a choice. The acoustical performance of these materials will be discussed in a later chapter.

Figure 8.14(b) illustrates a suggested plan for a large room, large audience system. To deliver the side surround reflections to all listeners the side walls are fitted with engineered diffusing surfaces (on the left) or curved surface diffusers (on the right). There are many opportunities for creative variations.

In all rooms there must be absorption and scattering, and fortunately in domestic rooms furniture, carpets, drapes often get us to where we need to be, about 0.3 to 0.4 s reverberation time. If not, more can be added. See Chapter 12 for guidance.

The addition of height channels in immersive systems changes things, but it depends on how the recording engineer chooses to use the additional channels. Here we are talking about baseline perceptions from listener-level channels. These are the most responsible for spaciousness and envelopment in multichannel music, movie music soundtracks, general "room" sound, audience sounds, etc. Getting them handled properly is the primary task. Overhead "objects" floating around will take care of themselves, as it is not likely that listeners will care much about their sound quality, even if they think they know what it should be.

8.5.6 Domestic Room Systems vs. Custom Home Theaters

Having experienced several in both categories I will say that there are no guarantees with either approach. I have been in expensive, visually lavish, custom theaters that had annoying problems, and in family room systems that delivered rewarding entertainment. Competence in choosing and using loudspeakers makes a big difference, as is no surprise to readers at this stage. Bass is a major factor in sound quality assessments (about 30%) and some installations simply dropped the ball, spending money on one monster subwoofer rather than on multiple smaller ones used intelligently. Bass boom is a constant annoyance. Acoustical treatment is a requirement in many installations, but 1-inch fabric-covered panels or 2-inch sculptured foam is not what is needed, however "cool" it looks. The job needs to be done well, whichever option is chosen.

Several factors influence one's choice. Cost is obviously one, as is the availability of space for a dedicated home theater. Another relates to lifestyle and how entertainment time fits into one's schedule, family activities, habits, and so on. Some people need dedicated rooms to get away from the kids, or to amuse the kids while they get on with their lives. Sometimes one needs an acoustically isolated space to enjoy music and movies at sound levels that normally would be an annoyance to others in the dwelling. For many, a TV on the living room wall, a soundbar, subwoofer and a couple of wireless surround loudspeakers is adequate.

With present technology it is surprising how good the best of these modest systems sound.

But the requirements for a “certified” home theater are anything but modest. In movies standardized playback sound levels apply to the dubbing stages in which soundtracks are mixed and cinemas in which they are reproduced (SMPTE RP200-2012). The measurements are made using a broadband pink noise test signal and a C-weighted sound pressure level measurement, with the meter set to slow response. The reference level is 85 dBC, for the screen channels, and 82 dBC for each of the surround channels. All these levels assume 20 dB headroom. This means that L, C, R screen channels must each be capable of 105 dBC, and each surround channel 102 dBC. These are not trivial sound pressure levels, so it is important to be able to predict in advance what is demanded of the loudspeakers and power amplifiers. High-end home theaters are often set up to meet cinema sound level requirements, even though many listeners find them excessive, both in home theaters and cinemas.

There is an entire industry built around custom home theaters. I know it well because for many years I gave tutorial lectures to large CEDIA audiences in different parts of the world that paid for the privilege (www.cedia.org). Nowadays, custom installers offer a host of home decoration, lighting and automation features, in addition to multiroom audio/video entertainment. It is a sophisticated business, but it is a business. Custom installers must make a profit which comes from them providing an aesthetic visual design, system design, room layout, seating, audio and video hardware, acoustical treatments that may include costly sound isolation, wired and wireless distribution systems, installation and calibration, a functional user interface, and on and on. Wealthy customers pay dearly for their services, and the results are status symbols that can look and sound impressive, and can entertain large families, including friends and neighbors.

Anyone having such thoughts should download CEDIA/CTA—RP22 (2023) “Immersive Audio Design—Recommended Practice”. It is a detailed exposition on the process, aimed at guiding custom installers, but also

providing much useful information for do-it-yourselfers. Full disclosure: Todd Welti was a contributor to the document. An earlier document, CEA/CEDIA CEB-22 (2009) is a simple guide for basic listener-level setups.

Fortunately, price and performance in audio and visual technologies are imperfectly correlated, so by choosing wisely, and knowing what one is doing, very high-quality entertainment can be obtained at moderate prices. One of the dominant cost savers is not having to achieve sustained calibrated cinema sound levels. Digital soundtracks in movies have liberated sound mixers to use more of the dynamic range than was possible in analog days. An expanded dynamic range is a good thing, but when it results in increased *average* not *peak* sound levels it can become a problem. Some movies, played at reference level in cinemas, have caused audiences to complain, even to walk out. There are reported instances where cinemas turn the volume down—some as much as 10 dB, half loudness, for some movies—to satisfy customers. A consequence of doing this is that the dialog is less intelligible. Nobody wins. A few films have such loud special effects that the dialog is unintelligible in places, played at any sound level. Recording engineers and producers creating the film know the script, cinemagoers don't.

There has been an ongoing argument among the artistic and technical people in SMPTE (Society of Motion Picture and Television Engineers) about reining in some of the loudness excesses, and there are no enforceable standards. Sensible movies use a large dynamic range, reserving the loudest parts for dramatic sound, gunshots, explosions, rocket launches, etc. Bass levels can be high without risking hearing, but high frequencies at comparable levels can be annoying and are a concern for hearing. Each film has its own loudness structure, so preferred playback volume levels will differ. With familiar films I sometimes modify the volume during the film, not wasting my hearing on gratuitous loud special effects, saving it for the worthy ones.

Complicating the situation greatly is the fact that a high percentage of our entertainment comes from streaming services, and they can differ

dramatically in sound level. The numerical volume setting on the playback system is therefore an unreliable guide to the listening level.

I have had home theater systems in my homes since about 1988 and I have rarely used them at reference playback level, except to show off. I value my hearing, and at my age (87) loud sounds can be uncomfortable, so one's personal factors enter the decision process. I have encountered a few people who seem to have a "deaf" wish, combining loud home theaters with loud hobbies; maybe the latter leads to the former. Chapter 16 discusses hearing conservation and problems—it is important to know that occupational hearing conservation guidelines, like OSHA (1997) and NIOSH (1998), are not designed to prevent hearing loss but to "conserve" enough so that at the end of a working life, one can carry on a meaningful conversation at 1 m distance in a quiet background. Hi-fi hearing is long gone. We need good hearing to benefit from all this knowledge and paraphernalia, and we get only one set of ears in a lifetime.

Increasing the overall volume is sometimes done to experience more bass. A subwoofer system can be controlled separately, and an overall bass tone control can be manipulated. A system with smooth, extended bass can be completely rewarding at less than cinema sound levels. If one is in a living situation with neighbors a wall or floor away, as many are and I am now, moderating sound levels is a requirement, especially as it is the low frequencies that travel throughout a structure. If the sound quality is high, sound levels can be lower—in my experience.

It isn't necessary to differentiate between stereo and multichannel rooms because so much multichannel material is based on stereo principles, even Atmos immersive music. The music soundtracks in movies are often stereo with added enveloping reflections in the surround channels. Music videos are frequently augmented stereo, and, of course, we can listen to basic stereo music. Any opinions one has about room treatment works for both, and good loudspeakers are good loudspeakers, for any program.

It may be worth revisiting Section 7.1 for a perspective on how consequential natural listening room reflections are, compared to the ones in recordings that we need to be fully satisfied. The more channels that are

delivering strong direct sounds the less the listening room itself matters. Put differently, above the transition/Schroeder frequency the listening room is most consequential in mono, less in stereo and even less in multichannel. Chapter 5 discusses this topic.

8.6 THE PRESENT AND POSSIBLE FUTURE OF MULTICHANNEL AUDIO

Time has passed, and it is no surprise that Dolby Atmos has become the leading system in the marketplace, but it is not clear that it is because of technical superiority, delivering the best audio experience, or simply marketing and brand name prominence—perhaps all three. The transition from stereo to ear-level multichannel was not entirely smooth, with old stereo habits very much in evidence. The appearance of elevation channels in immersive formats adds complications for content creators and for consumers. A certain amount of guesswork is involved, along with some trial and error, as these are not encode/decode algorithms. Literally, anything goes, so long as there is a customer for it.

Mixers may get to work with as many as 34 channels, but usually function with 7.1.4 or 9.1.6. Home systems start at 3.1.2 and proceed upward in channel count. This means that, even with the algorithmic intelligence of channel and object rendering schemes, it is not at all certain that listeners at home will hear what the recording engineer heard.

8.6.1 Upmixing, Downmixing and Rendering

Upmixing stereo recordings has a long history. It began with added loudspeakers at multiple locations, some delayed, some not, some with polarity inversion, and so on, all intended to add a persuasive spatial context to the front sound stage. One was sold by Dynaco as the QD-1 Quadapter. It delivered a sum of both front channels to a center-front loudspeaker and a difference signal to a center-rear loudspeaker. David Hafler (1970) was the inventor, and he proposed a quadraphonic recording system to complete the package. When multichannel systems became

available, enthusiasts saw an opportunity to improve upon the limited soundstage and spatial envelopment delivered by stereo. As has been mentioned, there are no rules for how stereo recordings are made, and there are no rules about how 2-channel stereo can be manipulated and/or augmented for upmixed delivery through 5 or 7 channels. This explains the wide range of opinions about how well stereo upmixing works. It is the combination of recording technique and upmixer that is being judged. With some combinations, the results are very pleasant. Sometimes the original stereo is preferred.

Much criticism of upmixing is related to how the frontal stereo soundstage is treated. Some upmixers generated unflattering renderings of the music soundstage, especially if too much center channel was employed as happened in some movie-based algorithms. In my experiences, David Griesinger's early versions of Lexicon Logic 7 with adjustable parameters were excellent for music upmixing, adding pleasant envelopment cues while minimally changing the stereo soundstage, but one parameter setting did not suit all recordings. However, the products containing those versions are long gone. Nothing since has been as satisfying, although I have found the Anthem upmixer and the Auro3D upmixer in some settings to be helpful without being offensive. It is postprocessing art, and there are more upmixers out there and possibly more to come.

It seems that current uncertainties relate to how the different configurations of multichannel and immersive audio are downmixed for delivery through different numbers of channels and different streaming services. Metadata is included in recordings describing loudspeaker and channel allocations and positions so that a rendering algorithm can adapt the recording to be compatible with the loudspeaker configuration in the playback system. It is something that technically inclined enthusiasts might be interested in, but in the end, customers are rarely informed about what they are listening to, and what compromises have been made in the delivery process.

8.6.2 A Radical Upmixing Alternative: Create Your Own Concert Hall

Of all the attempts to persuasively enhance the spatial illusions of stereo, one that I found very attractive was a byproduct of the Lexicon Acoustic Reverberation Enhancement System, LARES (Griesinger, 1991). This was an electroacoustic system of microphones and loudspeakers installed in performance spaces, all managed by a sophisticated digital processor. It allowed the reverberation characteristics of large auditoria and concert halls to be modified to be appropriate for different kinds of performances. Such systems have been very successful, and audiences are generally not aware of the invisible manipulations.

The version that got my attention was a version of LARES simplified for home listening rooms. I had one installed in the Harman demonstration home theater, and it was impressive. It consisted of four microphones and loudspeakers strategically located, all controlled by a digital processor. When one walked into the room, one was in a "hall". My favorite way of demonstrating the system was to play a multichannel classical recording with significant reverberation, and, while it was playing, I would gradually fade in the LARES enhancement. The change was never noticed, but when the music was paused the reverberation continued. Listener vocalizations, hand claps, etc. were then reverberated as if in a hall, not a home theater. It was impressive.

Stereo recordings played in the room with LARES enhancement were very gratifying and the benefits were available for popular music as well. The acoustical nature of the "hall" was adjustable. Unfortunately, at the time, the cost and complications were a deterrent for all but elaborate high-end home theaters, and the consumer version was not a commercial success. Nowadays, such enhancements could well be included as a feature of immersive systems as the loudspeakers are already there, and small microphones and digital processing are inexpensive. I have not heard of it being done.

8.6.3 Immersive Music Mixing Is Still Evolving

Regarding the delivery of immersive mixes. King and Theriault (2023) conclude:

> The industry is still in the early stages of providing immersive content to the consumer. Streaming platforms and binaural rendering formats are constantly evolving, as new research continues to inform on best practices within the immersive community. Mixing and mastering engineers are currently at a loss when it comes to delivering high-quality immersive music masters. There is no way to know with any confidence what the listening experience will be for the end user, the consumer. As binaural profiles and various encoding processes keep changing, immersive masters delivered even a few years ago risk sounding vastly different when streamed today and will most likely change again into the future.

On the topic of using the LFE channel in music mixes they say: "Heavy use of the LFE channel in immersive mixing and mastering may result in major differences in low frequency content between Dolby Atmos and Apple Spatial Audio binaural listening experiences."

A revealing investigation by Bourbon et al. (2023) analyzed the 7.1.4 channel-based render of 100 commercially released Atmos mixes hosted in a Hits in Spatial Audio playlist. They determined how much various channels were used for different components of mixes. There are several important findings, but one that stands out is that the majority are really embellished stereo mixes, with little use of the center channel: "Results presented suggest that many Atmos mixes are still primarily driven by the Left-Right speaker pair [i.e. stereo]. This is supported by regular identification of lead vocals, bass, and drums in the Left-Right speaker pair."

On the topic of LFE usage, they say that the sparse use of the LFE in Atmos mixes suggests "that engineers are concerned regarding translation between speaker and headphone-based playback."

Atmos appears to have influenced spatial rendering: "The identification of reverb, backing vocals, and guitars in the surround speakers suggests

space is used to widen the stereo image within an Atmos mix." They also note that:

> The rears are used more for envelopment purposes whilst the sides contain more rhythmic elements with clearly defined transients. This is further supported by low correlation between speaker pairs across all three objective measures. It is also clear that height is replacing depth compared with traditional stereo staging, with engineers situating reverb significantly above the listener compared with the frontal listening plane. Subjective analysis has uncovered frequent utilisation of height channels for reverb suggesting their primary use for providing envelopment and ambience. This is further supported by low loudness values and high brightness values.

These are still the early days of immersive sound, and only time will tell where it will go. In the meantime, it is probably best to relax and enjoy the music and movies, and not form lasting impressions and biases because, at present, even the creators of the mixes cannot be certain what we consumers are likely to hear in our home systems. Section 8.4.3 is an indicator.

8.6.4 There Is More to Come: Next Generation Audio

In its early years high fidelity, hi-fi, had the goal of replicating, as accurately as possible, live musical experiences. That goal has prevailed in certain quarters of audio over the decades, but the reality for most recordings is that they are abstract musical creations, carefully manipulated to deliver pleasing experiences through loudspeakers.

Traditionally, we take what we get, but enthusiasts have always made changes. Playback options like volume, loudness, tone and balance controls have been widely available to satisfy immediate needs and personal preferences. But everything moves on, and the current direction has altered to one of gratifying the wishes of the listener in ways never before possible.

Next Generation Audio systems, based on MPEG-H, promise much flexibility in choice of language, enhanced dialog, stereo-to-immersive perspectives with the ability to modify surround biases (Fraunhofer, 2024).

Also possible could be choices of phantom loudspeaker locations, flexible equalization, and more to be determined or invented. All of this will be superimposed on the upstream algorithmic or AI determinations of what sounds are delivered and in what processed form.

REFERENCES

Agrawal, S., Simon, A., Bech, S., Baerentsen, K. and Sorchhammer, S. (2019). "Defining Immersion: Literature Review and Implications for Research on Immersive Audiovisual Experiences", *Audio Eng. Soc., 147th Convention*, Paper 10275. ⏎

Bourbon, A., Watson, M., Stelmach, D., Champion, M., Chomal, A., Fenton, S. and Lee, H. (2023). "An Investigation into Immersive Contemporary Popular Music Mixing Practice Through Analysis of Released Music", *International Conference on Spatial & Immersive Audio, Audio Eng. Soc. Conference*, Paper 58. ⏎

CEA/CEDIA CEB-22 (2009). "Home Theater Recommended Practice—Audio Design". www.cedia.org ⏎

CEDIA/CTA—RP22 (2023). "Immersive Audio Design—Recommended Practice", Consumer Technology Association. www.cta.tech ⏎

Crompton, T. W. J. (1974). "The Subjective Performance of Various Quadraphonic Matrix Systems", *BBC Research Department Report No. BBC RD 1974/29*. www.bbc.co.uk/rd/pubs. ⏎

Dolby (2024). "Best Practices for Dolby Atmos Music Studios", September 5. https://dolby.my.salesforce.com/sfc/p/#700000009YuG/a/4u000000lFIV/YHgXgrJVKAcelcXkD.DnlRUBZU1CqEMGDKweLRpuu98. Accessed September 2024. ⏎

Fraunhofer (2024). "MPEG-H—the Next Generation Audio Standard for Broadcast and Streaming". www.iis.fraunhofer.de/en/ff/amm/broadcast-streaming/mpegh-audio-br-pt.html. Accessed June 2024. ⏎

Griesinger, D. (1991). "Improving Room Acoustics through Time-Variant Synthetic Reverberation", *Audio Eng. Soc. 90th Convention*, Paper 3014

⏎
Hafler, D. (1970). “A New Quadraphonic System”, *Audio*, July. ⏎
Holman, T. (1998). “The Basics of Bass Management”, *Surround Professional*, vol. 1, October. ⏎
ITU-R Recommendation BS.775-2 (2006). “Multichannel Stereophonic Sound System with and without Accompanying Picture”.
Julstrom, S. (1987). “A High-Performance Surround Sound Process for Home Video”, *J. Audio Eng. Soc*., vol. 35, pp. 536–549. ⏎
Keele, D. B. and Button, D. J. (2005). “Ground-plane Constant Beamwidth Transducer (CBT) Loudspeaker Circular-arc Line Arrays”, *119th Convention, Audio Eng. Soc*., Preprint 6594. ⏎
King, R. and Theriault, M. (2023). “LFE—Friend or Foe? The Pros and Cons of the Low Frequency Effect Channel”, *154th Convention, Audio Eng. Soc*., Paper 104. ⏎
NIOSH Occupational Noise Exposure (1998). www.cdc.gov/niosh/docs/98-126/pdfs/98-126.pdf ⏎
OSHA Occupational Noise Exposure. (1997). http://osha.gov/dts/osta/otm/new_noise/index.html ⏎
Ratliff, P. A. (1974). “Properties of Hearing Related to Quadraphonic Reproduction”, *BBC Research Department Report No*. BBC RD 1974/38. www.bbc.co.uk/rd/pubs. ⏎
SMPTE RP-200 (2012). “Relative and Absolute Sound Pressure Levels for Motion Picture Multichannel Sound Systems—Applicable for Photographic Film Audio, Digital Photographic Film Audio and D Cinema”. www.pub.smpte.org ⏎
Torick, Emil (1998). “Highlights in the History of Multichannel Sound”, *J. Audio Eng. Soc*., vol. 46, pp. 27–31. ⏎

CHAPTE
R 9

The Perceptual and Physical Dimensions of Sound

Floyd Toole

DOI: 10.4324/9781003477495-10

9.1 THE FREQUENCY DOMAIN

Figure 9.1 shows the horizontal axis of the frequency responses shown in this book. It displays frequency on a logarithmic scale because that is a convenient way to display several aspects of human perception. Only for special engineering purposes would a linear scale be used.

Figure 9.1 shows the horizontal axis of a standard frequency response graph, indicating the nominal high-frequency limit of 20 kHz, although only young non-noise-exposed listeners are likely to hear it. Beyond that is ultrasound. Conversely, 20 Hz is around the low-frequency limit for tonal perception—below it we tend to feel the vibration, the subsonic or infrasonic frequency range. A standard piano keyboard is shown, indicating the fundamental frequencies of the keys. An organ keyboard can cover the

entire audible frequency range, the lowest pedal notes being perceived mostly as pressure fluctuations. Other common musical instruments share portions of this range. This for the fundamental, pitch-determining frequencies. Harmonics and other overtones extend the frequency spectrum of most instruments to the upper limit of this scale, and beyond, into the ultrasonic frequency range.

In live performances, high-frequency sounds are normally much attenuated at listening locations because of atmospheric absorption, a phenomenon that progressively attenuates frequencies above about 1 to 2 kHz, and that increases with propagation distance (see Section 12.3.7). The loss of high frequencies with distance is itself a perceptual clue to the distance of *familiar* sounds. In concert venues, reflected sounds travel farther than direct sounds, thereby causing the reverberant sound field to rapidly lose high-frequency energy. Because reflected sounds dominate, live concerts are noticeably more "mellow" towards the rear of the concert hall. In amplified concerts, directional loudspeakers deliver mostly direct sound and are often equalized with compensating high-frequency boosts for more distant audience members, thereby bringing them perceptually closer to the performers. The same is true of recordings reproduced at home. Some listeners like the proximity, others wish for more "distance", but the very high frequencies are well communicated in recordings.

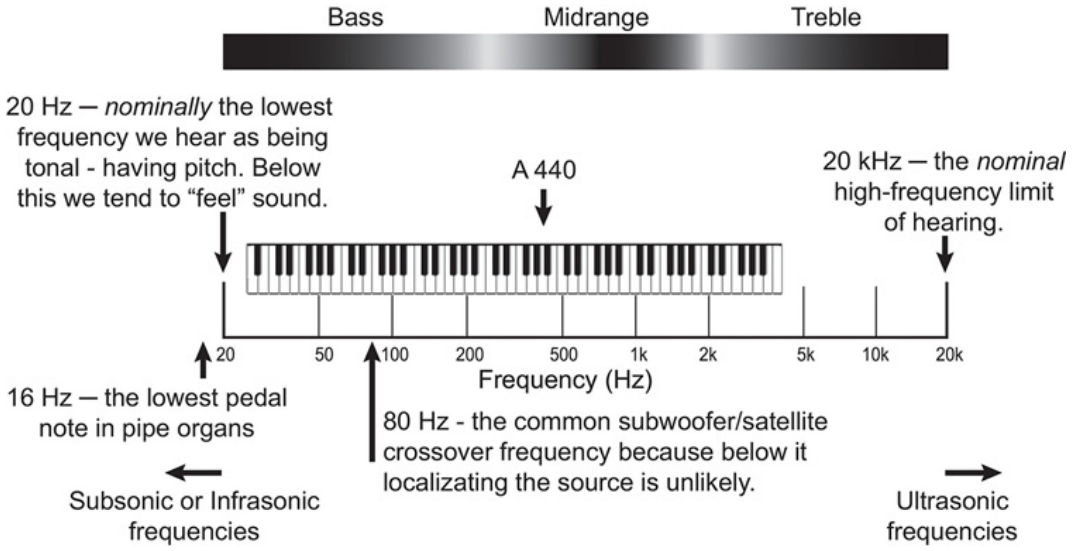


FIGURE 9.1 *The horizontal axis of typical 20 Hz to 20 kHz frequency-response data, the nominal hearing range.* ⏎

In close-microphone recordings these ultrasonic overtones can be captured if the microphones have the necessary bandwidth, which many do. Digital technology allows these frequencies to be recorded and delivered to consumers. This has caused some to believe that extending the recorded bandwidth has audible benefits; others regard it as a marketing feature. We will return to the topic later.

At the low-frequency end of the scale, things are complicated because humans can feel high-level low frequencies in addition to tonal perceptions. Pitch cues fade and vibration cues get stronger at very low frequencies. Feeling deep bass is pleasant, but the perception of pitch is the melody. The dual perceptions begin in the upper bass region, including kick drum "punch". Disproportionately boosted bass is common at rock concerts because it is dramatic to literally have your breath modulated by the kick drum and bass guitar—something not normally experienced at home. These can be "whole body" experiences. Fortunately, low frequencies are not high-risk sounds for hearing loss, but the rest of the accompanying spectrum definitely is hazardous. Sound-attenuating musicians' earplugs or loudness-limiting headphones are recommended.

In perceptual terms, the correlate of frequency is pitch. It is not a perfect relationship because most musical instruments radiate complex spectra and waveforms, not pure tones. The judgment of pitch is therefore somewhat dependent on the relationship of the fundamental frequency and the overtones. The common example of this is the piano in which, because of insufficiently flexible strings, not all the overtones are harmonically related to the fundamental. Consequently, "stretched" tuning is necessary to make the highest notes perceptually acceptable, and in smaller pianos, with shorter strings, this is required at low frequencies as well. Pitch is also dependent on sound level, with low notes sounding slightly lower and high notes slightly higher at increased volumes. The consequence of this to sound reproduction is obvious: the music itself changes with sound level, so

a somewhat realistic playback level is a requirement if “realism” in all its manifestations is the expectation.

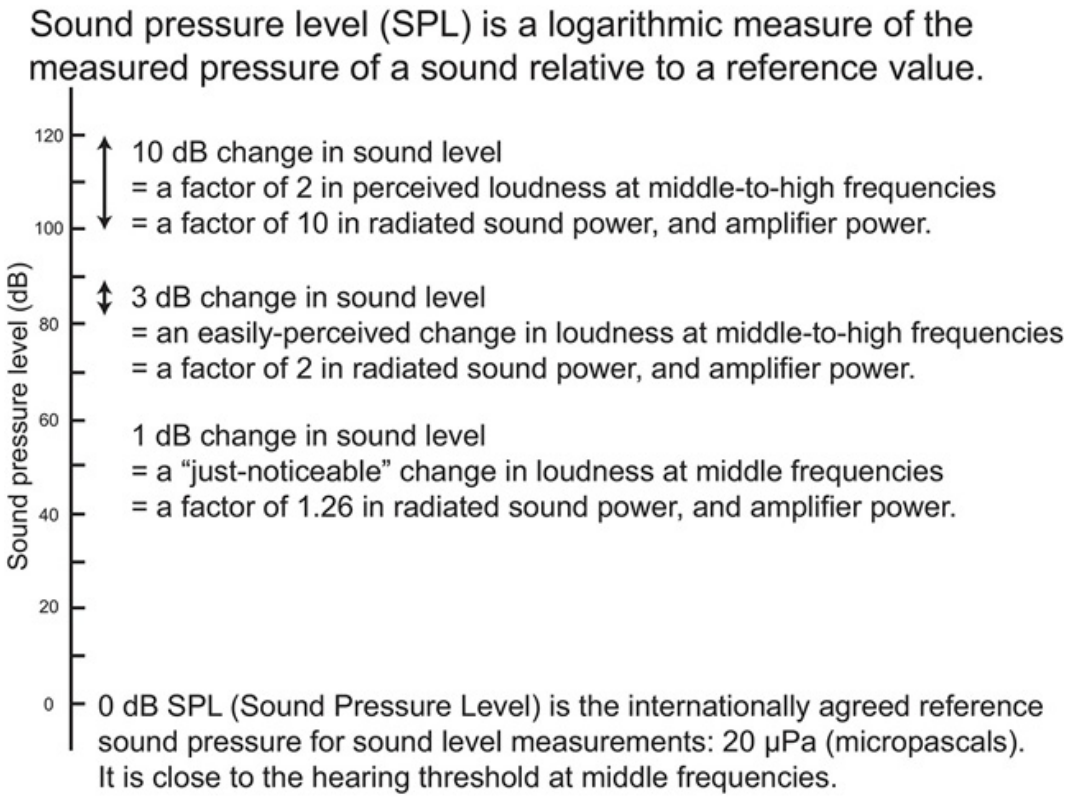


FIGURE 9.2 *The Sound Pressure Level (SPL) measurement scale references sounds to a standard sound pressure of 20 micropascals, which defines 0 dB SPL. The indicated 10, 3 and 1 dB changes in level are not related to any specific SPL. These differences apply to broadband program material, not narrow band or bandpass filtered material. Emphasis over a portion of the frequency range, as in the bass response or resonances, yields different results.* ⏎

9.2 THE AMPLITUDE DOMAIN

Humans respond to sounds ranging over a sound pressure range of about 1 million to 1. When interpreted as sound intensity (energy flow per unit area) it is an even more impressive 1 trillion to 1. This enormous range extends

from sounds that are just audible at middle frequencies to those that generate discomfort or pain at the higher limit. Because such large numbers are impractical to work with, a logarithmic scale of sound pressure level is used to describe measured and audible acoustic events. The units are decibels (dB). Figure 9.2 introduces the scale.

9.2.1 Absolute and Relative Sound Levels

Sound Pressure Level, or SPL, is used when discussing *absolute* levels of different sounds relative to an internationally agreed reference level. It assumes unweighted measurements—all frequencies are equally represented. In loudspeaker measurements it is used because it shows the absolute sound pressure level at a standardized distance (1 m) (measurements are usually made at 2 m for far-field accuracy, and increased by 6 dB for 1 m) for a standardized voltage input (2.83 V) to the loudspeaker. This defines the loudspeaker's sensitivity. The frequency response graphs shown in the book show only a portion of the audible SPL range for clarity. Most of the graphs in this book use the same scaling to make it easy for readers to make comparisons. Many published graphs use different scales, compressing or expanding features of interest, to exaggerate or to diminish aspects of performance. This can be done for extra clarification, or for outright deception. Pay attention to the numbers on the vertical scale and determine what it is that was measured. It is not uncommon for graphs to be shown without any indication of what was measured.

The decibel scale can also be used to compare any two sounds, disregarding the standardized reference 0 dB. In such cases it is correct to use "sound level" or "relative sound level" instead of sound pressure level as the notation on the vertical axis if the graph displaying the differences in sound levels.

Figure 9.2 shows visualizations of significant differences in sound level. A 10 dB level change is perceived as approximately double or half loudness for mid- and high-frequency sounds. The difference in power output from amplifiers or loudspeakers is a substantial factor of 10. A 3 dB change is a

nicely audible difference, a kind of “turn the volume up a little, please” magnitude of difference. The change in power is a factor of two, which is not trivial. A 1 dB change is a detectable level change in everyday listening and the power level change is small, a factor of 1.26.

9.2.2 Different Interpretations of Sound Pressure Levels

Figure 9.3 shows the approximate sound levels of different sound sources, and other relevant information. The sound sources identified with relative sound levels in the left-hand column have been collected over many years; some are true, unweighted sound pressure levels (SPLs), but many were measured with unspecified (probably A-weighted) instruments, using unspecified time averaging. A-weighted measures have a strong mid-to-upper-frequency bias (see Figure 10.19), giving them greater relevance to hearing risk assessment. Measuring distance is often unknown, and many of the sources identified are extremely variable, so these are very crude estimates. However, it is evident that, when considered in the light of potential damage to hearing, the opportunities are numerous in everyday life, and unavoidable in the military and recreational shooting, unless very effective hearing protection is worn.

Unweighted sound pressure levels (SPL) and A-weighted sound levels are not equivalent unless the sound source is narrow band, falling within the A-weighting contour. For example: if a sound has a strong low-frequency component it will be ignored by an A-weighted measurement, but will influence or may even dominate an SPL measurement.

The column pertaining to hearing damage looks alarming, and it is. Individuals are highly variable in their susceptibility to hearing loss, so population averages are used. The occupational hearing conservation guidelines are important for those who work in noisy occupations, 5 days a week for a working lifetime. Applying them to occasional exposures is incorrect but often done. It is assumed that these guides prevent hearing loss. That is not the case. They permit hearing loss to occur, aiming only to conserve enough at the end of a working life to permit carrying on a generally intelligible conversation at a distance of 1 m in the quiet.

Appreciating the subtleties of live or reproduced sounds would be a pleasure long gone. To workplace noise exposures must be added non-occupational noise exposures: rock concerts, shooting, motorcycles, lawn mowers, power tools, automobile air bag deployment, etc. all of which contribute to the lifelong accumulation of hearing damage. It is estimated that sounds above about 75 dB can contribute, slowly but surely, to the accumulated lifetime hearing loss. Such sound levels are unavoidable in life, which likely contributes to presbycusis, age-related-hearing loss, but there are several other factors involved with aging bodies.

> Hearing that is considered "normal" by your audiologist, whose criteria are based only on speech intelligibility, may already exhibit significant losses when it comes to appreciating or evaluating the many spatial and timbral dimensions of music and movies. It is known that seemingly innocent hearing losses degrade one's ability to form reliable sound quality judgments.

The column at the far right of Figure 9.3 is an estimate of the amplifier power needed to deliver various sound levels in a typical room. Obviously, this is fundamentally dependent on loudspeaker sensitivity, how many are operating, the frequencies being radiated, and the acoustical properties of the room. In spite of these uncertainties, the *relative* quantities are correct—the overall vertical scale may slide up or down by a few dB. This is shown to give an appreciation for the reality that a 3 dB change in sound level requires a doubling or halving of power, and a 10 dB change, translates into ten times the power. When one is talking about high sound levels from domestic loudspeakers having typically moderate sensitivity (85–90 dB SPL at 2.83 V at 1 m), one is also talking about very large amounts of power. Fortunately, for most people it is not a requirement.

Taking full advantage of wide system dynamics requires some serious power from amplifiers and loudspeakers, as is shown in Figure 9.3. In reality, we make do with (and enjoy) much less dynamic range than is

possible. The reason that occasional loud sounds emerge from modest audio systems is that we can tolerate up to 6 dB of clean amplifier clipping of peaks (no ringing or instability) with remarkably little complaint; 6 dB is a factor of 4 in power (Voishvillo, 2006). Those experiments used music. In movies *very* loud events are mostly sound effects with no real "fidelity" issues, since the original could possibly be the sound of an elephant trumpeting, equalized, and played at quarter-speed, backwards.

9.3 AMPLITUDE AND FREQUENCY TOGETHER: FREQUENCY RESPONSE

Many important measured quantities, electrical and acoustical, are displayed as an amplitude/magnitude vs. frequency curve. The one of greatest importance in audio is the amplitude of electrical signal or sound level vs. frequency, or as it is commonly known: the frequency response. It is the principal determinant of sound quality. Knowing what is "correct" in examining a frequency response curve is important, so Figure 9.4 identifies some of the key factors that can be revealed in these curves. The example is a loudspeaker, driven with the standard test signal level of 2.83 V, and the sound pressure levels shown are corrected to the standardized measurement distance of 1 m, so it is possible to see the *sensitivity* of the loudspeaker as a function of frequency (dB SPL at 2.83 V at 1 m). Curves (a) and (b) were measured on the reference axis of the loudspeaker, which would normally be directed at the listener, providing the direct sound in a room. However, loudspeakers radiate sound in all directions, and that sound, less any that is absorbed, eventually reaches listeners as reflections, so many more measurements must be made, as indicated in the family of curves shown in (c). All of this is explained in detail in Chapter 5.

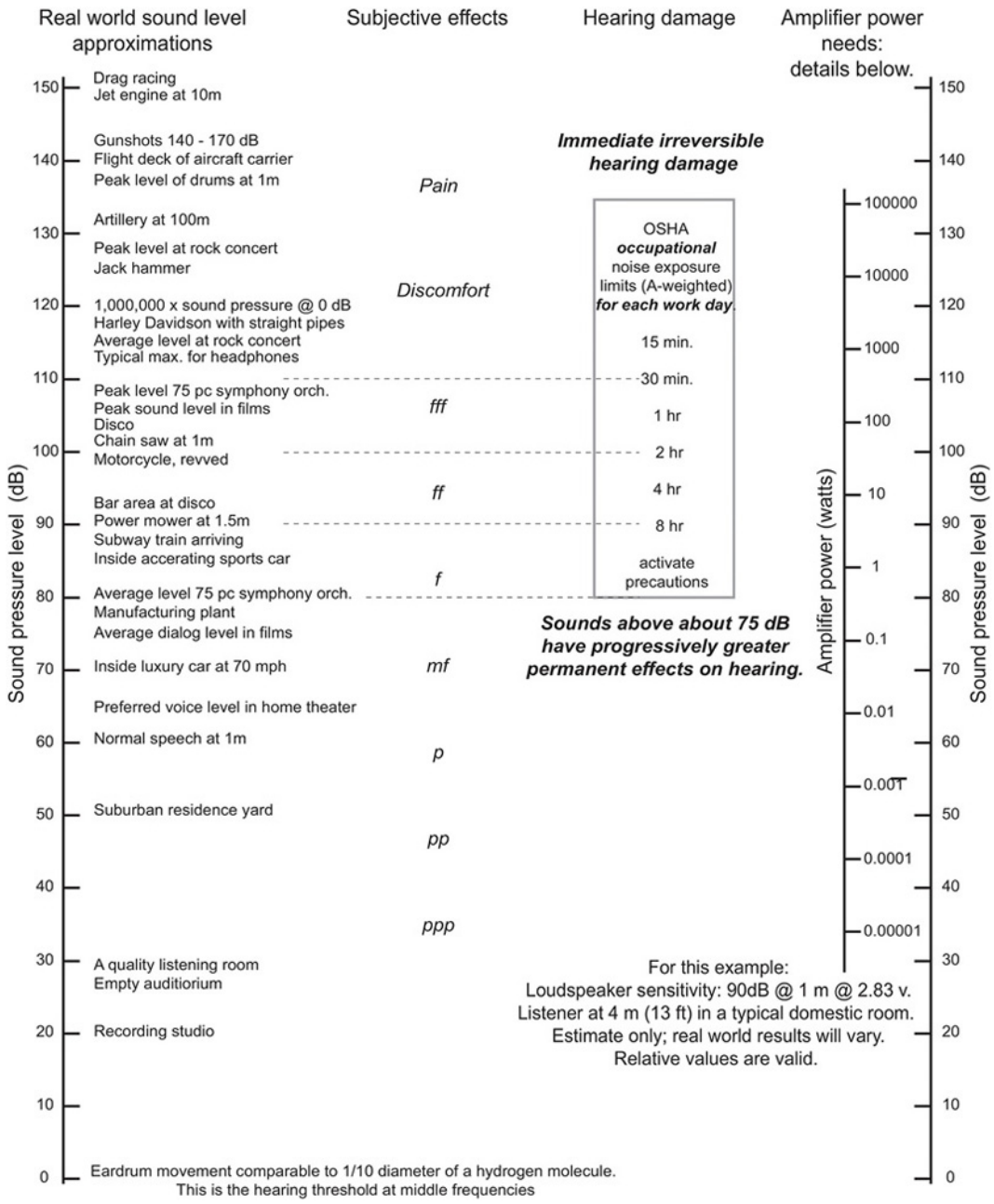


FIGURE 9.3 *Examples of various sound levels that we may be exposed to in various situations are shown on the left. Although the left scale is labeled Sound Pressure Level (SPL), some of the examples were measured with A-weighting, and the OSHA recommendations are A-weighted measurements. All the sound level estimates on the left are approximate because of uncertainties in how the measurements were made and normal acoustical variations in real-world circumstances. It is intended*

to be a rough guide. In the middle are descriptions of a few subjective effects and considerations related to preserving one's hearing. On the right are idealized estimates of amplifier power required to produce these sound pressure levels in typical domestic rooms using relatively sensitive loudspeakers. These are single frequency levels, meaning that broadband sounds will be louder, and will vary with content. The relative levels are still valid. ⏎

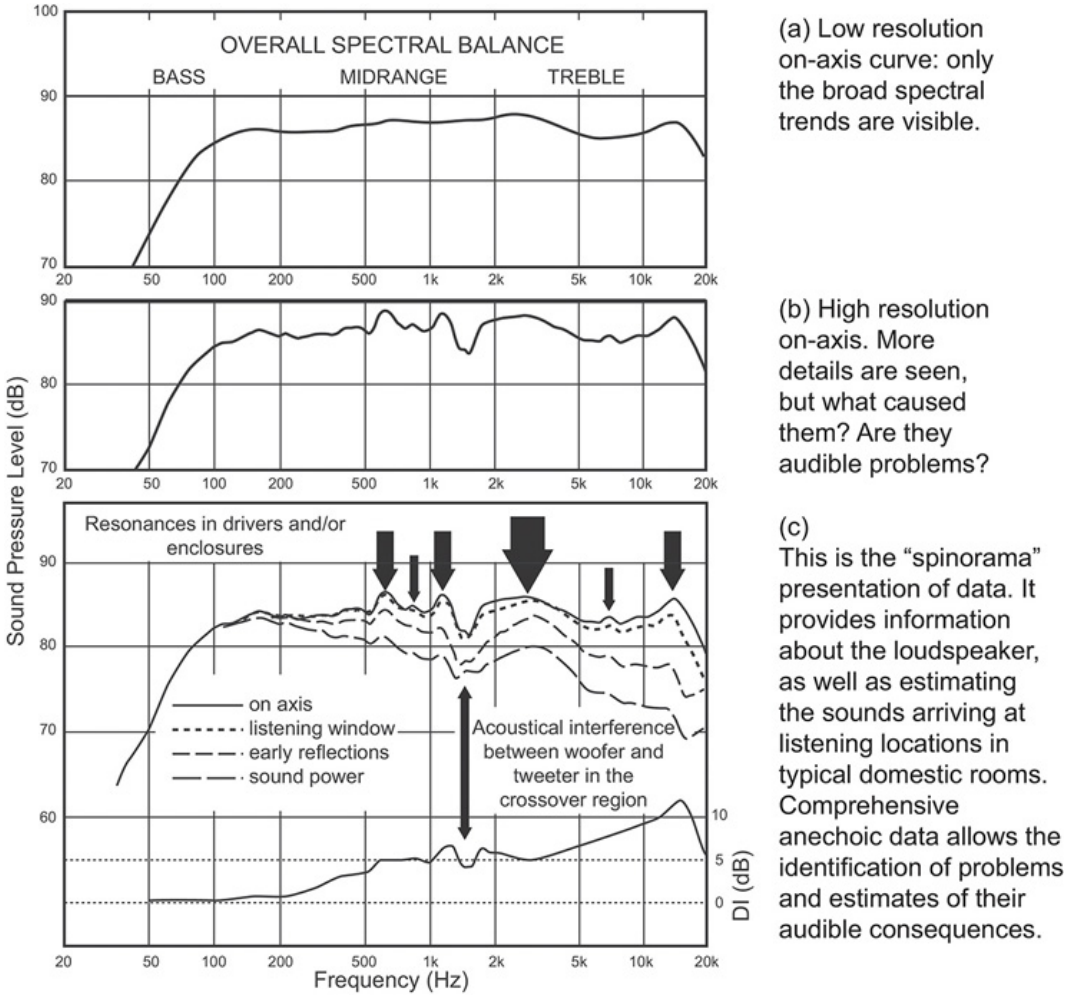


FIGURE 9.4 *An illustration of some information that can be revealed in frequency-response curves. In these examples the sound pressure level at 1 m is shown when the loudspeaker is driven by a constant 2.83 V. This is an industry standard format.*

Figure 9.4(a) shows a spectrally smoothed curve revealing upper and lower frequency limits and general trends in sound output. (b) is a measurement with higher frequency resolution that reveals other possibly audible timbral degradations. (c) shows the family of curves informally called the spinorama that has proved to be useful in identifying the mechanisms underlying the features in frequency-response curves, and quantifying them in a manner that permits assessment of their potential audibility in rooms. One frequency-response curve cannot describe the performance of a loudspeaker. ⏎

Figure 9.4(a) provides a starting point for speculation about the sound quality of this loudspeaker. This is an "artistically smoothed" interpretation that often appears in product specifications and sometimes in product reviews. It is widely assumed in marketing circles that the public is not ready to see something close to the truth. It can be seen that the bass output is rapidly attenuated below about 80 Hz, meaning that a subwoofer would be needed for satisfying bass, and 80 Hz is a very convenient crossover frequency for bass management systems (Section 8.2.2). The sensitivity of the loudspeaker over the important middle portion of the frequency range is somewhat frequency dependent, which is why sensitivity specifications are usually based on the sound output averaged over a range of frequencies, such as 300 Hz to 3 kHz. The slight upward trend in output from about 100 Hz to 3 kHz suggests that the sound may be somewhat thin, bright or hard depending on the program. Perceived "brightness" is often associated with frequencies much lower than is commonly thought. The depression at higher frequencies deprives the listener of balanced instrumental overtones (see Figure 9.1) possibly leading to a loss of openness and articulation, perhaps even a hint of dullness.

Figure 9.4(b) shows more details, and possible problems, but additional information is needed to identify what they are with any certainty. The spinorama shown in Figure 9.4(c) is an example of such data—measures of sounds radiated in other directions. The specifics of these curves have been

discussed in Section 2.7. One of the most important features of this family of curves is its ability to reveal the presence of resonances, the most annoying contributor to a lack of neutrality in sound quality; see Sections 1.5.1 and 5.3. Most of what we hear in typical rooms is reflected sound, and the lower curves reveal increasing amounts of reflected sounds. It is a property of transducer resonances that they tend to be widely radiated into rooms, which means that the bumps associated with resonances will be evident in all the curves, and that they are likely to be audible in rooms. Because of their wide dispersion, resonances may not be evident in the directivity index (DI).

Resonances that occur within the normal operating ranges of transducers are *minimum-phase* phenomena, meaning that a bump in an amplitude response is sufficient evidence to predict that there will be a localized phase anomaly and ringing in the time domain. Section 5.3 discusses this in detail. It turns out that the shape and height of the amplitude response bump is the most reliable indicator of audibility. This and the fact that these resonances can be attenuated by precise equalization mean that these anechoic data are very important. As is explained in Section 5.3, the most audible resonances are the low-Q (i.e. spectrally broad) ones, such as that seen here around 3 kHz. Attenuating just this one spectral feature would greatly improve the sound quality of this loudspeaker. The resonance near the high-frequency limit is unwanted but in normal program is not likely to be annoying. Those in the middle frequencies are more obvious because they affect the timbre of voices and many musical instruments.

Interested readers can return to Section 5.3 for a detailed review of resonances.

Acoustical interference can arise in the interactions between transducers radiating simultaneously in crossover regions, and also from direct sound interfering with sound diffracted from edges of enclosures, and the like. These peak and/or dip shapes change with direction, sometimes averaging out over a range of angles, and at other times, as in the present example, being sustained over a range of angles but diminishing in the ultimate sound power spatial integration. This is why it is visible in the directivity index,

the difference between the listening window and normalized sound power curves. Because it is different in different reflected sounds, most acoustical interference effects are less audible than resonances.

Good engineering design can reduce or eliminate any of these problems, but it is important first to know what they are. It is obvious that a simple product specification like ±3 dB 80 Hz to 18 kHz contains almost no useful information for consumers, especially in the common situation where there is no tolerance and no description of how or where it was measured. Section 2.9.1 provides a real-world example.

9.4 AMPLITUDE AND FREQUENCY TOGETHER: EQUAL-LOUDNESS CONTOURS

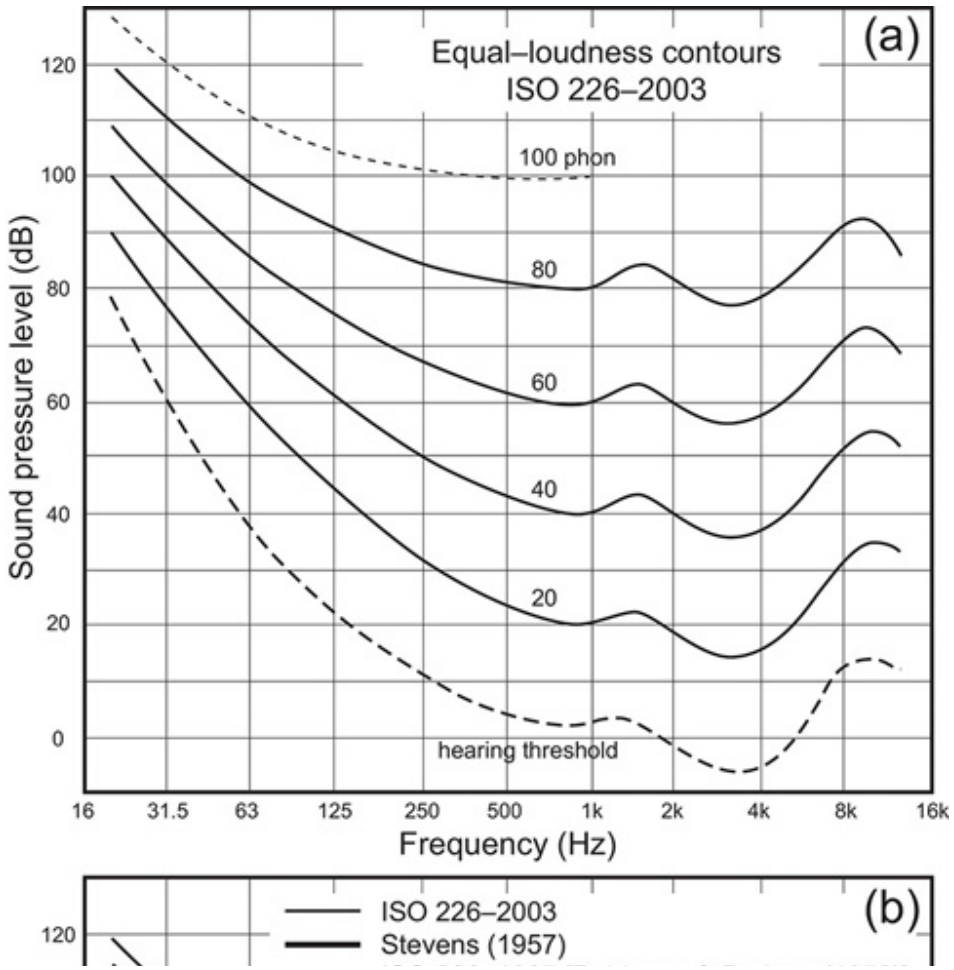
Loudness is the perceptual correlate of sound level. But it is also dependent on frequency, incident angle, the duration of the signal and its temporal envelope. It is one of many psychoacoustic relationships that defy simple descriptions. Yet, one must start somewhere, and so most investigators focused on pure tones in acoustically simple circumstances, like headphone or anechoic listening.

Fletcher and Munson (1933) were among the first to evaluate the sound levels at which we judged different frequencies to be equal in loudness. Starting with a reference pure tone at 1 kHz, other tones at different frequencies were adjusted until they appeared to be equally loud. With enough of these measurements it was possible to draw a contour of equal loudness identified by the sound pressure level of the reference tone, but called *phons* rather than decibels, to denote its subjective basis. For each SPL of the 1 kHz reference tone, curves were derived, collectively called the equal-loudness contours. Fletcher and Munson used headphones, and there has been concern about the way they were calibrated, so while we give them credit for their pioneering work, their data are no longer trusted.

Several more recent studies, notably the work of Robinson and Dadson (1956), used pure tones presented to listeners in an anechoic chamber; their results were the ISO standard for several years. Even then, errors were

identified in the contours at low frequencies. Pollack (1952) evaluated a set of equal-loudness contours for bands of noise, but the circumstances of the listening were not well specified. Stevens (1957) moved closer to typical listening conditions by employing fractional-octave bands of steady-state sounds in reflective spaces, which he used as a basis for calculating the loudness of complex sounds (Stevens, 1961). The Stevens curves shown in Figure 9.5 are from Toole (1973) scaled to the present format.

Figure 9.5(a) shows a sample from the current ISO standard (ISO 226, 2003), the result of an international collaboration initiated to address errors in the 1987 edition of the standard that was based on Robinson and Dadson (1956) data. Figure 9.5(b) shows a comparison of it and contours from four earlier studies showing how very different they are in shape. These are obviously *not* hard engineering data. The curves are averages over many listeners and there were large inter-subject differences.



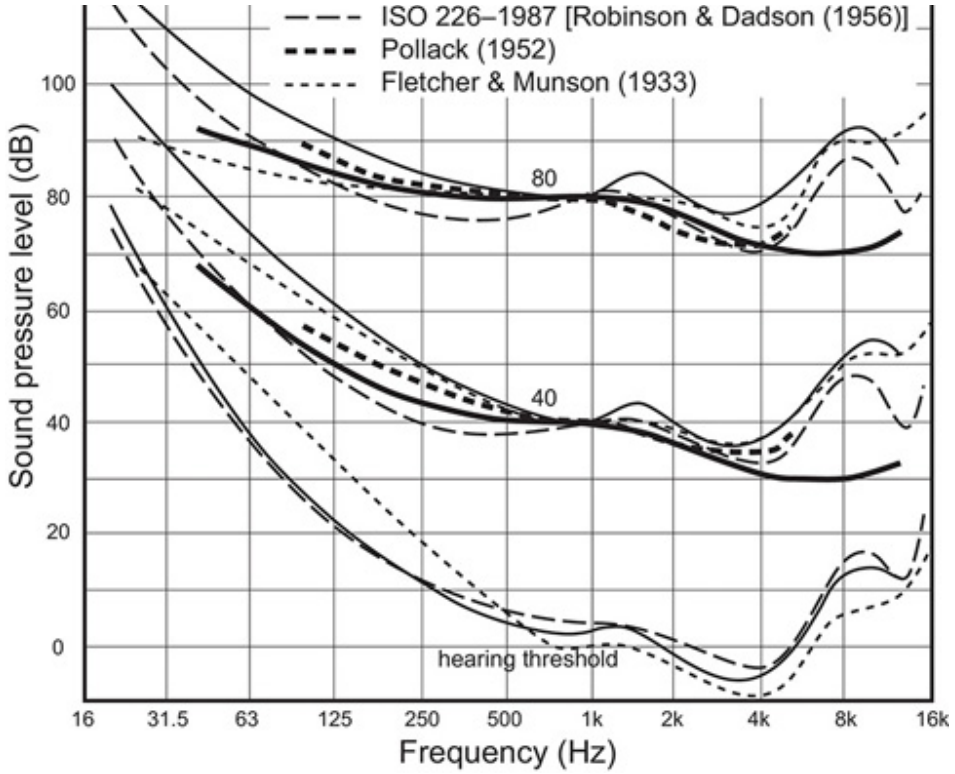


FIGURE 9.5 *(a) Curves of equal loudness for pure tones auditioned in an anechoic space according to the most recent internationally agreed standard: ISO 226 (2003). Each curve is created with reference to a pure tone at 1 kHz, and is identified by the sound pressure level at that frequency, but expressed in "phons". (b) A comparison of the current standard with three other determinations of equal-loudness contours. Equal loudness contours are from ISO 226 (2003). © ISO. This material is reproduced from ISO 226 with permission of the American National Standards Institute (ANSI) on behalf of the International Organization for Standardization (ISO). No part of this material may be copied or reproduced in any form, electronic retrieval system or otherwise or made available on the Internet, a public network, by satellite or otherwise without the prior written consent of the ANSI. Copies of this standard may be purchased from ANSI, 25 West 43rd St., New York, NY 10036, http://webstore.ansi.org. All other curves are reproduced with permissions from the respective sources.*

Nevertheless, the curves all share two important characteristics:

- Our sensitivity to sound reduces dramatically as frequency is reduced. Bass sounds must be much higher in sound level to be perceived to be as loud as mid- and high-frequency sounds.
- The curves progressively crowd together at lower frequencies, meaning that smaller changes in sound level (dB) are required to yield the same change in perceived loudness (phons). This is responsible for the very familiar and easily audible rapid growth and decline of loudness at bass frequencies when the overall volume is changed. The contours are very similar at frequencies above about 500 Hz, with a relatively constant separation. At these frequencies a change of 10 dB in sound level corresponds to a change of 10 phons—which is a roughly a factor of two in perceived loudness. However, at lower frequencies the curves move closer to each other. At 63 Hz the relationship is about 7 dB for 10 phons, and at 31.5 Hz it is about 5 dB for double loudness. Small changes in sound level at bass frequencies are audibly consequential something experienced in program material with some regularity—the circle of confusion (Figure 1.8).

9.4.1 Loudness Controls: Do They Work?

"Loudness" controls have been in audio equipment for decades. The attractive premise is that as the volume is reduced, automatic adjustments to frequency response compensate for the disproportionately reduced bass. For no obvious reason, some of them also boost the treble. This is wrong. It is also wrong to think that a simple variable bass boost can compensate for all components in a program that has dynamic range; loud portions need only slight bass boost, while quieter portions will need much more boosting. If the loudness compensation addresses the low-level portions, the high-level portions will have greatly excessive bass and so on.

The equal-loudness contours describe a characteristic of our hearing systems that is always there and unchanged whether we are at a live musical

performance, having a conversation, or listening to music or movies at home. They are part of the human listener, not something that needs to be "corrected" for, as some have mistakenly believed. They are typically measured using pure tones at different frequencies that are compared to a reference tone at 1 kHz and adjusted to be equally loud. It is a simple process using simple sounds, and the real question is the extent to which these curves relate to the complex continuous spectra of music. It is an issue to which there is no definitive answer.

Nevertheless, they cannot be ignored in sound reproduction. For a sound reproducing system to accurately portray what was heard in a live performance it must have a uniformly *flat* frequency response and reproduce the sounds at or close to the original sound levels, so that our built-in loudness processes operate on the sound in the same way. This is where complications set in, because that is rarely the case. Even if our sound systems have flat frequency responses, the playback sound levels are not likely to be what they were at live performances or in recording control rooms where important artistic decisions were made. For recreational listening, sound levels are almost always lower, often much lower, than the "original" sound.

To understand this issue it is necessary not to pay attention to the shapes of the curves, but instead to the *differences* in their shapes at different sound levels. Two things happen as playback sound levels are reduced:

1. The bass frequencies are disproportionately reduced in apparent loudness compared to the rest of the spectrum.
2. Progressively more very low frequencies will fall below the hearing threshold, becoming completely inaudible.

Olive (2004a, 2004b) found that in the overall ratings of sound quality, bass accounted for about 30% of the total factor weighting. It is important to get it right, and it cannot be right if some of it is missing.

Figure 9.6 shows results of an exercise performed by the author over 50 years ago (Toole, 1973) to explain the problem: if the loudness of music is reduced by, say 10 or 20 phons, what new frequency response must the playback system have in order that:

a. the timbre or spectral balance of the original will be preserved in the less loud reproduction, and
b. any frequency audible in the original will be audible in the less-loud reproduction.

Condition (a) cannot be perfectly satisfied because music has dynamic range and different components of the program at different levels will require different compensations. Music has large dynamic range, *ppp* to *fff*, as shown in Figure 9.3. The ear is a non-linear device generating distortions, pitch shifts and more that are sound-level dependent. These cannot be replaced.

Condition (b) is a concession to human nature—we don't want to lose any of the music.

Figure 9.6 shows results for three loudness reductions. A 10 phon reduction from a high "reference" level (a) is a substantial change (approximately half loudness), but one that is not at all uncommon. Most people would still consider this to be "foreground" listening. A 20 phon reduction (b) yields "background" music, and a 30 phon reduction (c) is for "ambiance". It is immediately obvious that no single loudness compensation curve can work for music that has significant dynamic range. The low-frequency boosts are substantial to maintain audibility of low-level sounds, and these would render high-level sounds grossly bass heavy. Clearly condition (a) will not be satisfied, but it may be worth considering a compromise solution, and some musical realities may moderate what is needed. First, low bass sounds are most often in high-level musical passages because even in live performances the sound levels must be high in order for these sounds to be audible—Figure 9.5(a). It therefore seems

reasonable to focus the compromise compensation on the higher-level components, such as the 60 and 80 dB SPL portions.

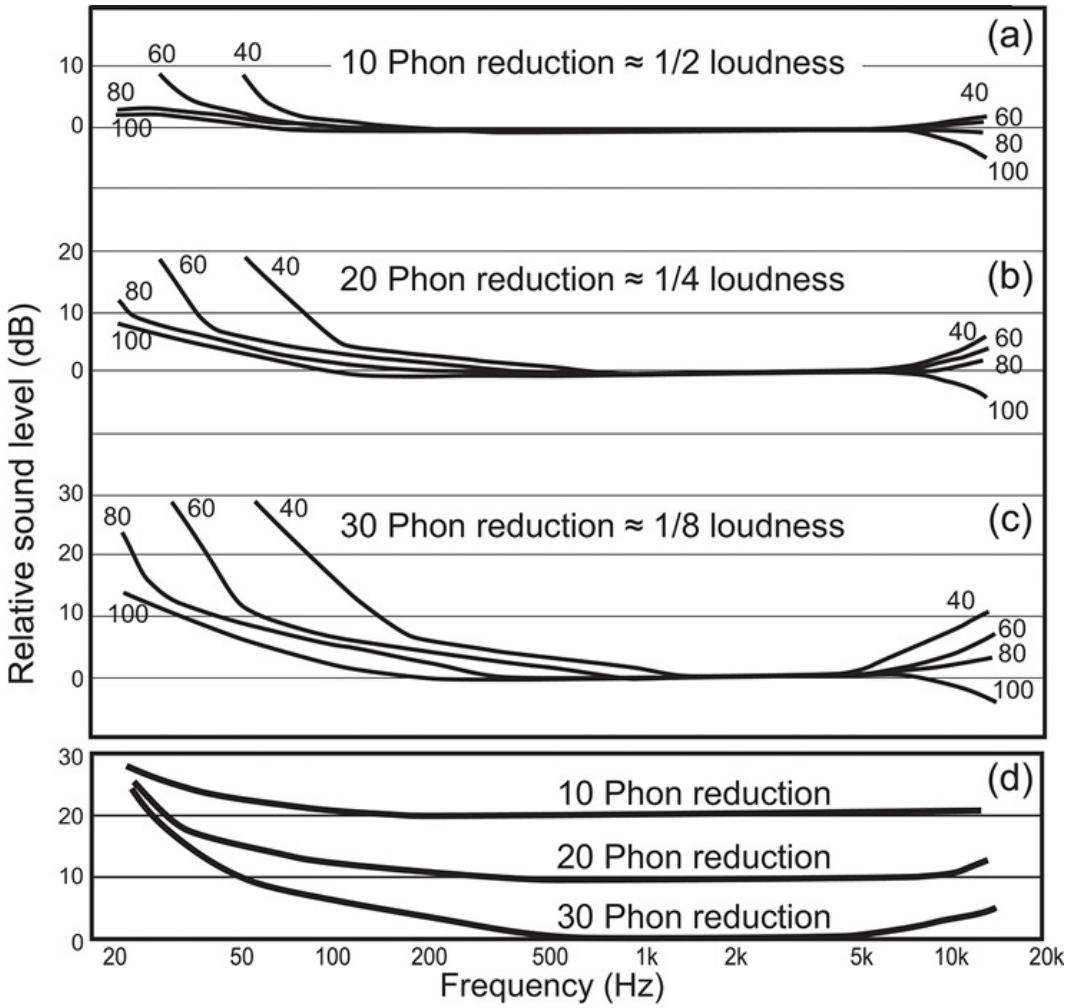


FIGURE 9.6 *The frequency-response changes needed to maintain the apparent spectral balance and audibility of sounds having flat spectra at original sound pressure levels of 40, 60, 80 and 100 dB SPL when the loudness level is reduced by (a) 10 phons, (b) 20 phons and (c) 30 phons. Derived from Robinson and Dadson (1956) equal-loudness contours. (d) shows compromise curves drawn between the 60 and 80 dB predictions. Data shown in (a) to (d) are from Toole (1973).* ⏎

Figure 9.6(d) shows the result of plotting compromise curves between the 60 and 80 dB curves in each of (a), (b) and (c). These look quite practical. It is important to note that no great changes are required at high frequencies. Many loudness compensation devices over the years have significantly boosted both bass and treble as sound level was reduced. With such controls, sound quality is diminished—a result of misinterpreting the equal-loudness contours. With 10 or 20 phon reductions essentially no high-frequency compensation is needed.

This esoteric discussion might be interesting, but it all is an approximation based on an approximation (the obsolete Robinson and Dadson contours), based on experiments done with pure tones, not broadband music. If there is a message, it is that attempts at loudness compensation for playback at less than original sound levels are destined to be imperfect, but they might contribute to improved entertainment simply by keeping some low bass sounds from being less audible or disappearing altogether.

9.4.2 Loudness, Tone and Tilt Controls—A Summary

Listening at sound levels below the "original" or intended level causes timbral changes and a progressive loss of low bass unless corrective measures are taken. Unfortunately, because music has dynamic range, the corrections cannot be correct for all components of the original sounds. Strongly amplitude compressed music would respond more favorably to such correction, though, meaning that much modern music would certainly qualify. It is possible to choose a compromise loudness correction, catering to the needs of the louder portions of programs. When this is done, the substantial loudness compensations are at low frequencies—essentially it is a bass tone control.

Movies are mixed and reproduced in cinemas at sound levels that are standardized. Home theaters are often calibrated to meet a similar sound level for 0 dB volume control setting. This is a situation where a calibrated loudness compensation could function somewhat predicably, but the dynamic range problem would still exist. The reality is that playback at 0

dB in a home theater is commonly regarded as being uncomfortably loud and volume settings are often reduced by 10 dB or more. An interesting current issue with movie sound in cinemas is that many customers find some of them too loud and complain to the management. The problem has reached a point where some jurisdictions have talked about placing legal limits on the playback sound levels in cinemas. As mentioned earlier, some cinema owners have reduced the playback sound levels by as much as 10 dB—half loudness! For those few blockbuster films where directors have insisted on sustained, penetrating, loud sound (compensation for an anemic plot?) this is understandable. But the dialog is reduced along with the annoyingly loud passages, and intelligibility suffers. This is a serious situation because the "art" is being compromised. Some sounds, gunshots, explosions, etc. *must* be loud to be persuasive and to convey appropriate dramatic punctuation. However, prolonged loud sounds that seem unnecessary for the events portrayed on the screen are simply tiresome and annoying. Good taste is the answer. Hearing loss is not a key issue for casual moviegoers, although the sound mixers should be concerned.

Home theater installers have long appreciated that many customers are pleased if the system is equalized to have a bass boost—sometimes a significant one. As has been said, excessive bass is a "forgivable sin" for many. Music programs are totally uncalibrated in record or playback levels, requiring music lovers to listen and adjust a bass control until the sound is pleasantly balanced spectrally. The curves in Figure 9.6 do not call for a treble control, although there may be other reasons to use one, and loudness compensators that automatically boost the treble along with the bass are to be avoided—they are ill-conceived.

It has become a part of high-end audiophile folklore that tone controls damage the "purity" of sound, and some equipment proudly has no such controls. The underlying assumption must be that recordings are flawless, but from common experience this is simply not the case. Some recordings have insufficient bass, others too much. Some are too bright, others too dull, and there are other audible maladies.

It is relevant to consider what professional recording engineers think is good spectral balance, because their opinions get incorporated into the recordings they create. Mäkivirta and Anet (2001) measured room curves at the recording engineer's location in 250 control rooms, using functionally similar Genelec loudspeakers (see Figure 9.7). Although the median curve is commendably flat, the variations that exist in these control rooms can be large, especially at low frequencies. Recording engineers who work in these circumstances demonstrate the truth in the "circle of confusion" (Figure 1.8). They cannot be certain of hearing the same sounds when moving among different recording studios, much less anticipate what consumers are hearing with any certainty.

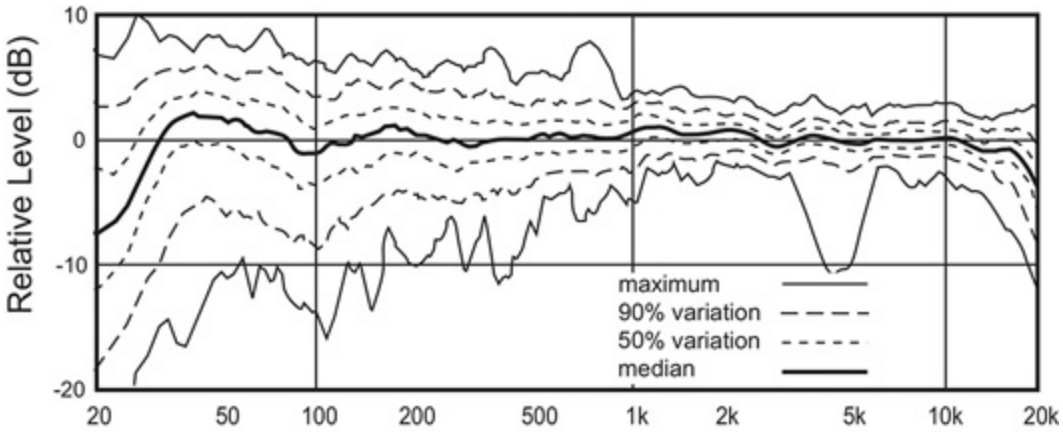


FIGURE 9.7 *Room curves measured in 250 recording control rooms, at the engineer's location, using functionally similar Genelec loudspeakers. Scaled by author to match the book standard. Mäkivirta and Anet (2001).* ⏎

Many recording engineers insist on using their own recordings when setting up new mixing locations. This can perpetuate the characteristics of the original site. There is no opportunity for improvement, of moving to more timbrally neutral territory. The widespread distrust of measurements in general is a problem, but if room curves are the *only* measurements being looked at, there is an even greater problem. Professional audio has some notably unprofessional aspects especially when it comes to relationships

between subjectivity, the art they create, and objectivity. The devices used to create the art and by customers to appreciate it need to be technically measured for subjective comments on the art to be truly meaningful.

The fact that we may choose to listen at lower than “reference” sound levels at times means that a bass tone control is a requirement, not an option, in achieving rewarding listening experiences. Easily accessed tone controls allow listeners to find their own compromise settings in dealing with the characteristics of their audio systems and rooms, and in real time, following the needs of individual programs.

Peter Walker was the founder of Quad, and I had the pleasure of knowing him. He gave us what have long been respected as relatively neutral sounding electrostatic loudspeakers, implying that no “equalization” was needed. Nevertheless, he incorporated a “tilt” tone control in his Quad 34 Preamplifier (ca. 1982). It was basically bass and treble controls linked so that as one increased, the other decreased. Because the entire spectrum was “tilted”, the audible effects were easily heard, so the boost/cut was limited to about 3 dB. There was also a separate low bass control. The explanation was simple: not all rooms are the same and personal tastes vary, so find the spectral balance you most like. He went on to explain that recordings are not consistent in their bass/treble balance, excessive brightness being a common problem at the time, and this could be compensated for. Then for classical works one may wish to assume a different perspective on the orchestra than the recording engineer provided, and a tilt control gave the listener some ability to change this. Altogether it was a very sensible addition to an audio system, adding value—and participation—to the listening experience. Much later, Lexicon incorporated bass, treble and tilt controls in its surround processors. Others may also have done so. It is a good idea. The author has used them extensively over the years.

In terms of how sensitive we are to spectral tilts, there is evidence that we can detect slopes of about 0.1 to 0.2 dB/octave, which translates into a 1 to 2 dB tilt from 20 Hz to 20 kHz (Kommamura and Mori, 1983). Such a spectral error, if small, is likely to be quite benign to sound quality, and subject to adaptation: we simply would get used to it. However, if large

enough, it causes those of us who have them to reach for the bass, treble or tilt control to restore a more pleasant spectral balance.

Not all playback electronics have these controls, and some of those that do require entering a menu and/or waiting for a digital reconfiguration, which are disincentives to using them. “Old-fashioned” knobs or the real-time digital equivalents are user friendly. High-end “audiophile” electronics that deliberately avoid tone controls do customers a disservice, and customers who avoid them are missing a feature that can occasionally improve sound at zero cost.

9.5 THE BOUNDARIES OF WHAT WE CAN HEAR

Figure 9.8 summarizes some aspects of sound and hearing that collectively bear on sound reproduction. The bottom of the display begins with a shaded area representing sounds that are below the statistically determined threshold of audibility. The short wavelengths at high frequencies make it difficult to be sure of the sound pressure level at the eardrum so conventional audiometric measurements stop at 8 kHz. However, extended or high-frequency audiometry is available, allowing measurements to 20 kHz. A quick review of some results confirm that both age and noise exposure can dramatically elevate very high-frequency thresholds. The younger population exposed to extended headphone listening seems to be at risk. The dot-dashed curve on the right side is a crude attempt to visualize results of high-frequency audiometry, as there is no “official” statistical threshold statement yet. There is considerable individual variation among subjects as well as some issues related to calibrations of the measurements. Also at the bottom is an example of the noise floor of a good microphone, indicating that it is possible to capture most of the smallest sounds the ears can hear.

Moving up the display from the hearing threshold, some examples of NC (Noise Criteria) curves used for evaluating background noise levels. These represent background sound spectra that might find their way into recordings, or that might be present during their playback. Here the curves

have been lowered by 4.8 dB to make the octave-band NC curves compatible with the 1/3-octave filtered music and car interior noise. The true NC curves are shown in Figure 9.8.

In the middle of the display is a collection of background noise measurements made inside cars at highway speeds which, when combined with simultaneous masking, explains why under these circumstances music loses much of its bass, timbral subtlety and spatial envelopment. Only in the parking lot or in stop-and-go traffic can good car audio systems reveal their true excellence. Quiet cars are highly desirable. It also explains why car audio systems are often balanced to have rather more bass than would be usual in the home, and why elaborate ones incorporate bass boost activated according to road speed and/or interior noise. Surround channel levels would similarly benefit from that kind of automated adjustment.

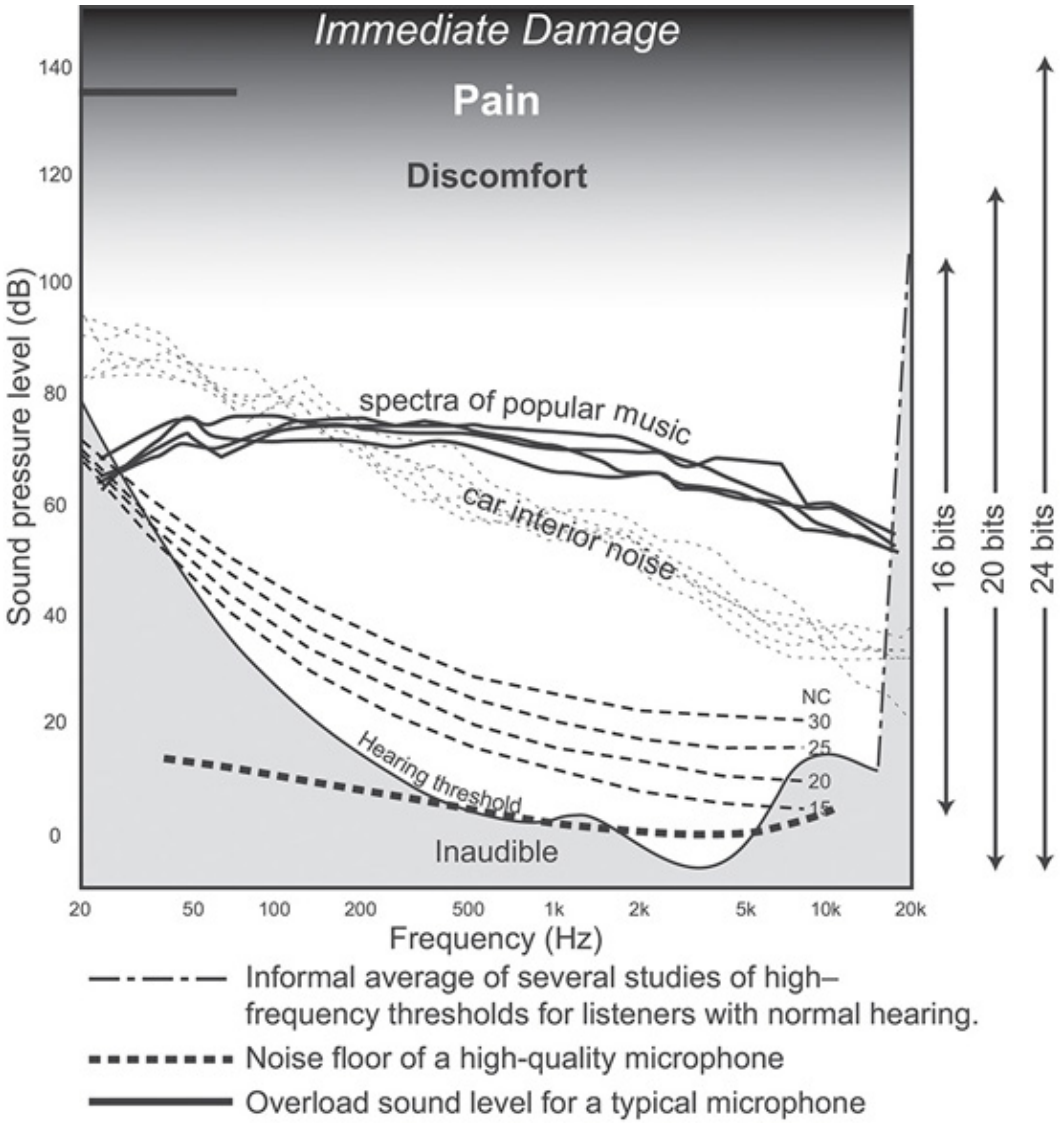


FIGURE 9.8 *A simplified two-dimensional display of what we can record, store and hear. At the bottom is the hearing threshold, and an example of the noise floor of a good microphone. Above this are some examples of typical background sounds in recording and listening environments, including cars at highway speeds, to high sound levels where things become uncomfortable and then permanently damaging. Running across the middle of the white area representing the useful listening region are examples of long-term average spectra of popular music at foreground listening levels (from Olive, 1994). Peak levels will be higher. On*

the right are shown estimated dynamic ranges of perfect digital record/playback systems. ⏎

All of this is made directly relevant by the set of music spectra slicing through the middle of the display. These come from measurements done by Olive (1994) and represent four programs exhibiting unusually flat and extended spectra. These are long-term average spectra, which ignore whatever dynamic range the program had. The background car noise exceeds these spectra below about 200 Hz, explaining why car audio systems need large bass boosts at highway speeds.

At the top of Figure 9.8 are shown regions of discomfort and pain and the absolute barrier of instantaneous severe hearing loss or deafness. These are the reasons why turning up the volume is not the answer to elevated background noise levels. It becomes tiresome, and ultimately damaging.

On the right are shown the dynamic ranges of some digital record/playback systems, showing that the conventional 16-bit CD, if well used, has the potential to be a satisfactory delivery format for most program material. But 20 bits would be better, allowing leeway for some imperfection in the process. And 24-bit systems permit the encoding of everything from "I can't hear it" through to "I'll never hear again". It makes sense in recording studios during the creative process, but it is excessive as a delivery format. Taking full advantage of wide system dynamics requires some serious power from amplifiers and loudspeakers, as is shown in Figure 9.3.

9.5.1 What Is Acceptable Background Noise?

Over the years various criteria have evolved aimed at enabling acoustical designers to set goals for background noise levels that are acceptable. Most targeted commercial spaces, and the criterion was speech: how well could normal speech be understood against a background of HVAC and other building noises? Although simple weighted sound level meter measurements can provide a rough guide, it is usual to start with an octave-band analysis of the noise spectrum, and then to compare this spectrum to

one of several criterion curves purporting to describe the acceptability of the background noise for different specific purposes.

Measures driven by such concerns remain at the core of the popular criteria used in setting acceptable background sound levels in listening and recording spaces even though speech interference is not a problem. In those situations one is likely to be more concerned with whether the noises are annoying or pleasant. In North America, the NC curves are widely applied to define acceptable background noise levels for audio environments. "[the NC curves] are intended to be octave band noise levels that just permit satisfactory speech communication without being annoying"; so said the creator of these curves, Beranek, as reported in Tocci (2000). Warnock (1985) states:

> NC contours are not ideal background spectra to be sought after in rooms to guarantee occupant satisfaction, but are primarily a method of rating noise level. There is no generally accepted method of rating the subjective acceptability of the spectral and time-varying characteristics of ambient sounds. In fact, ambient noise having exactly an NC spectrum is likely to be described as both rumbly and hissy, and will probably cause some annoyance.

This does not sound like an objective for an expensive studio facility or recreational listening space.

Tocci (2000) and Broner (2004) provide lucid surveys of the numerous families of contours proposed by acousticians for various purposes. If the sound quality of the background noise is important—and this should be the case in listening environments—there is agreement that it is worth considering something other than the traditional NC contours where the shape of the spectrum is not evaluated, only the penetration of the highest point in the spectrum into the family of curves (the tangency criterion). For audio purposes the noise spectra should be as smooth as possible. The most recent NC contours are shown in Figure 9.9.

NC criteria levels that are generally considered acceptable for various sound reproduction venues are listed below. They are approximations. Obviously lower levels are preferred if they can be achieved.

Recording studios—distant microphone pickup	NC-5 to NC-10
Recording and broadcast studios—close microphone pickup	NC-15 to NC-25
Concert halls and other live performance spaces	NC-15 to NC-20
Home theaters, suburban homes	NC-15 to NC-25
Movie theaters	NC-30 to NC 35
Urban residences and apartments	NC-30 to NC-40

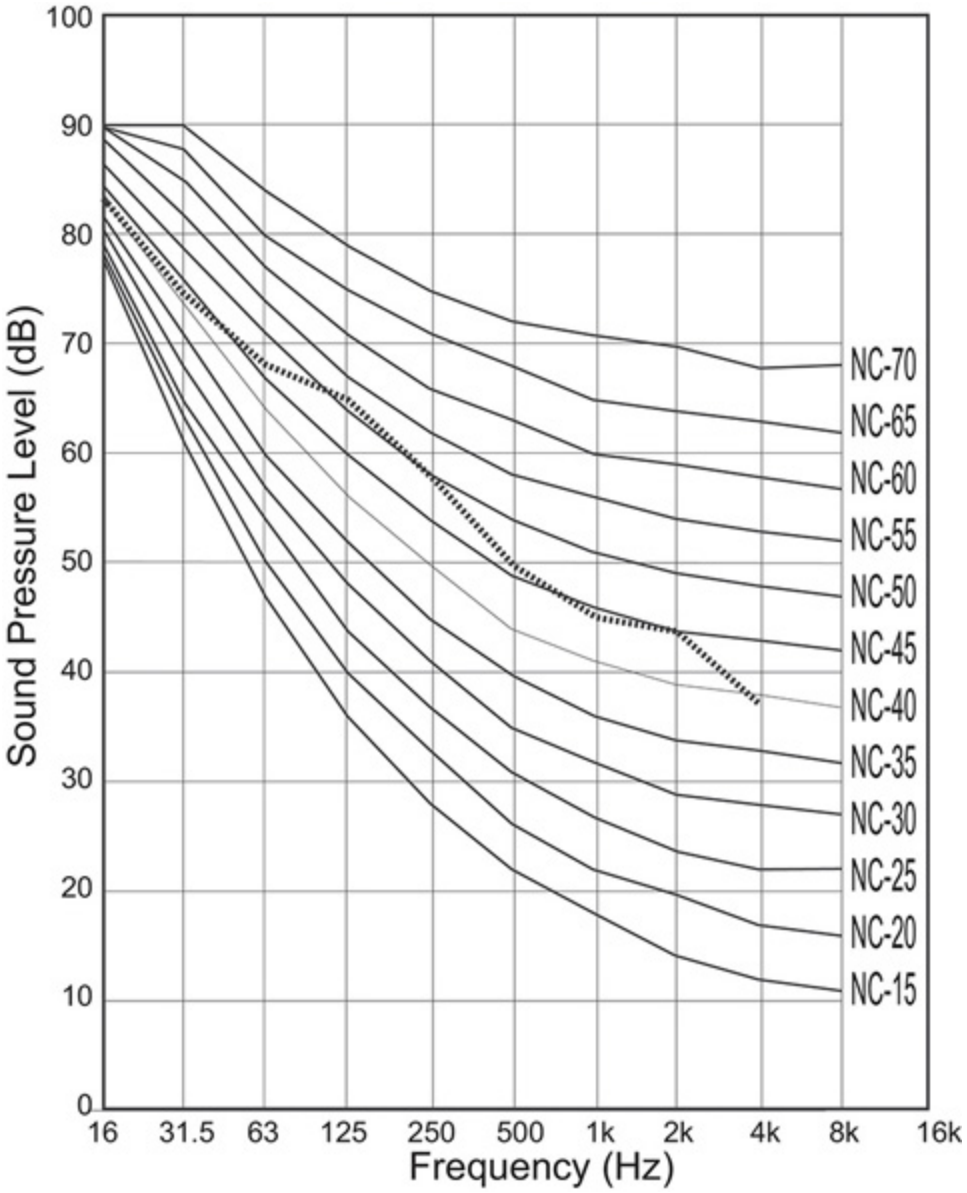


FIGURE 9.9 *The new NC contours showing a measured spectrum (the dotted curve) registering NC-51 according to the tangency criterion. The NC rating of a spectrum is designated as the value of the highest NC curve "touched" by the measured octave-band spectrum. Reprinted from ANSI/ASA S12.2:2008 American National Standard on Criteria for Evaluating Room Noise, © 2008, with the permission of the Acoustical Society of America, 1305 Walt Whitman Road, Suite 300, Melville, NY 11747.* ⏎

Commercial or multiple-dwelling buildings with massive central HVAC systems present the greatest problems. When making distant microphone recordings in some venues it may be necessary to shut down the HVAC to reduce the background noise. In these cases two NC ratings should be provided, one with HVAC and one without.

After a detailed investigation, the ultimate goal for recording studios was concluded to be as close to the hearing threshold as possible (Cohen and Fielder, 1992). Everything else leaves a residual background that inevitably will have a character of its own, and which might be captured in recordings and reproduced in quiet listening rooms. Figure 9.10 shows this perspective on background noise, which includes some real-world data on home listening rooms. Clearly, with rare exceptions, any sound isolation in home theaters is most likely to be needed to keep good sounds in, rather than bad sounds out.

A very quiet room is very impressive, and customers like to demonstrate the silence, whether or not it matters once the movie or music is under way. Home theaters that are acoustically isolated from the rest of the house to prevent the escape of theatrical dynamics almost always end up being very quiet—the sound transmission loss works both ways.

9.6 LINEAR DISTORTIONS: AMPLITUDE, PHASE AND POLARITY

Audio devices should ideally yield outputs that are perfect replicas of the signals that went in, or perfectly scaled versions of those signals if amplification is involved. Transducers (microphones and loudspeakers) convert energy from one form to another, and again the requirement is that the output is a perfectly scaled transduction of the input signal (e.g. voltage in and sound level out). The success of this is evaluated by measurements comparing inputs and outputs, looking for discrepancies—distortions.

Linear distortions are those deviations from perfection that are independent of sound level. The shape of a waveform (amplitude vs. time) —the music—may be linearly distorted, but that distortion is the same at all

sound levels. Linear distortions can change the relative levels (amplitude) or timing (phase) of frequency components within the complex music signal, and thereby alter the shape of the waveform. The combination of amplitude and phase as a function of frequency is commonly known as the transfer function that constitutes the Fourier transform of the impulse response; they are equivalent information. Humans are very sensitive to variations in amplitude vs. frequency, and very insensitive to phase shifts. In other words, we do not "hear" waveforms per se, but we are very sensitive to their spectral content.

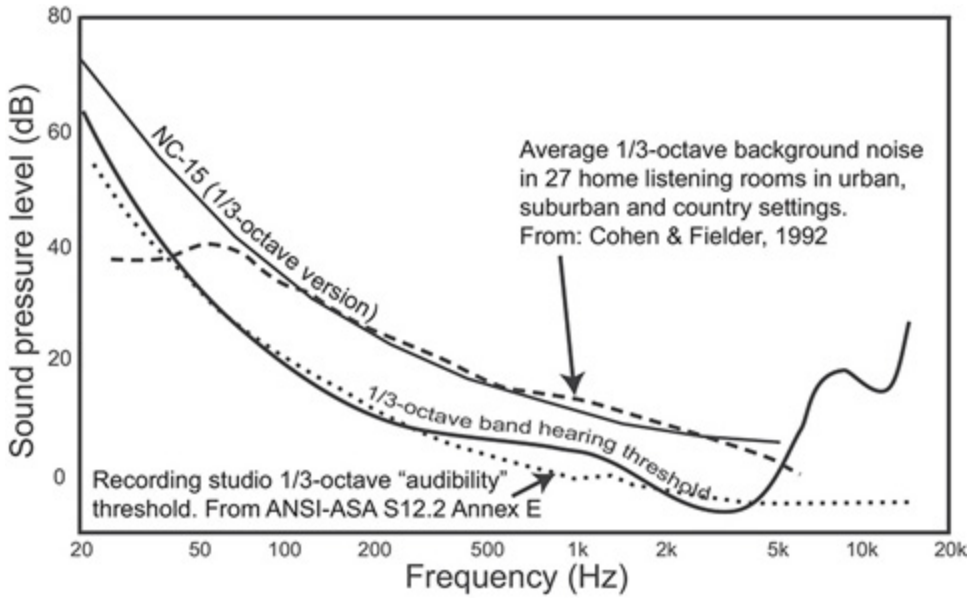


FIGURE 9.10 *An interesting comparison of data originating in the Cohen and Fielder (1992) paper. The bottom "audibility" curve was motivated by this paper and underlines the need to apply more restrictive background noise requirements to audio recording and listening environments than has been traditional. Fortunately, many homes are adequately quiet. The NC and ANSI-ASA curves are reproduced with permission of the Acoustical Society of America.* ⏎

If the waveform changes as a function of signal or sound level those changes are non-linear distortions—these non-linear mechanisms *add*

frequencies to the signal that were not there originally and are especially objectionable if they are high enough to be audible.

The amplitude vs. frequency curve, which is known as the frequency response, is the single most important measurable parameter of any audio device, and the general objective is a smooth flat curve extending over the audible frequency range. However, that goal has yet to be universally achieved in our imperfect physical world. Consequently, especially in loudspeakers, we must learn how to evaluate the audible consequences of different kinds of deviation from the ideal—the audibility of tilts, peaks, dips, bumps and wiggles. In the following we will look closely at the audibility of deviations from what might be considered to be target performances.

9.6.1 Spectral Deviations From Flat and Smooth: Resonances

The dominant spectral features that translate into an audible characteristic of sound quality are those associated with resonances. Section 1.5.1 announced this fact, and Section 5.3 probes deeply into the amplitude and time domain measures of audible resonances. It turns out that amplitude response deviations are the most reliable indicators of the audibility of resonances, and that, with low-Q, broadband, resonances, very small deviations are audible. These broadband events extend to tone-control manipulations, broadband parametric and shelving equalization. The narrow, sharp spikes associated with high-Q resonances and prolonged ringing are, when listening to normal programs, much less disturbing because they are less often and less energetically stimulated. Because transducers are minimum-phase devices, evidence of a peak in the frequency response is proof of ringing in the time domain. It is not necessary to measure in the time domain to know that it happens. However, technology permits us to see both amplitude and time domain information together.

9.6.2 Resonances Viewed in Amplitude, Frequency and Time: Waterfalls

Simple resonances exhibit their presence in the frequency and time domains. Things get more complicated with room resonances, soon to be discussed, because what you hear also depends greatly on the room, and where you and the sound sources are located in it—"space" becomes a factor.

Waterfall diagrams are very photogenic, and intellectually appealing because they combine the three domains of amplitude, frequency and time. The problem is that the audio print and internet environment is awash with waterfall diagrams that are inadequately described. They may show what is claimed; they may not. These displays have been used to inform, to impress and, unfortunately, also to deceive. The observer needs some help in order to know what the display is revealing. This provocative introduction needs explanation.

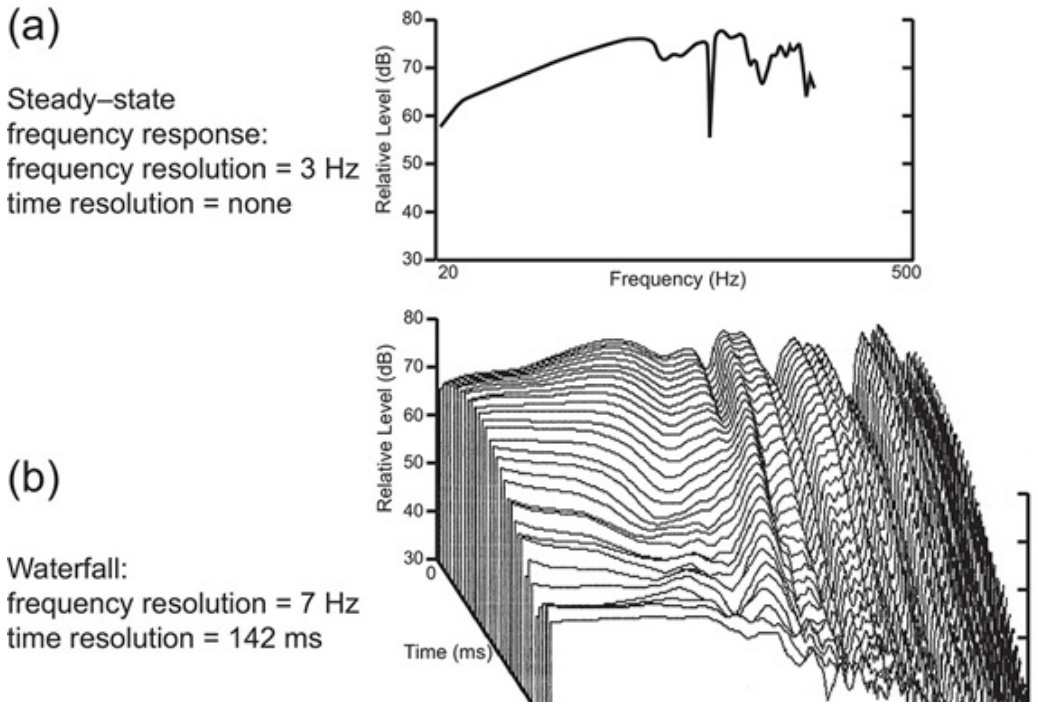
Ideally, we want to be able to see high-resolution data in the frequency and time domains simultaneously. Unfortunately, this is not always possible; in general, you can have one or the other, not both. Figure 9.11 illustrates some examples of the consequences of making different choices in frequency and time resolutions.

Figure 9.11(a) shows a very high resolution (about 3 Hz) steady-state frequency response. Many measuring devices, including some relatively inexpensive laptop-based packages, can generate this kind of data. It has excellent detail in the frequency domain, but it reveals nothing *directly* about what is happening in the time domain. However, because we know that low-frequency room resonances generally behave in a minimum-phase manner, we know that, if there are no prominent peaks protruding above the average spectrum level, there will not be prominent ringing in the time domain. It is this *indirect*, inferential, knowledge that permits us to confidently use frequency responses as a primary source of information about room behavior at low frequencies. Figure 9.11(b), (c) and (d) show the *same* situation displayed in time and frequency, as a waterfall, but employing different resolutions.

In Figure 9.11(b) the frequency response at the back, near time = 0, is closely similar to the curve in (a), because a narrow 7 Hz bandwidth was

chosen. However, the successive curves comprising the waterfall all look very much like it because in achieving high resolution in the frequency domain, the time-domain resolution has been sacrificed; it is 142 ms. Within the total 500 ms decay that is shown, nothing much changes because every successive curve moving forward is an average of events within a 142 ms window. The first several curves after t = 0 are almost identical, and those following simply drop smoothly in amplitude. It looks as though this room has uncontrolled ringing at many frequencies, but this is wrong. The data are misleading. This kind of pattern is frequently seen in audio discussions.

In Figure 9.11(c) the frequency resolution has been reduced; now it has an effective resolution of about 14.3 Hz. The top frequency response is smoother, revealing less detail but, because of the improved 70 ms time resolution, we now can see some detailed changes in the decay curves. All the curves decay more rapidly and there is evidence of what seems to be a frequency shift in one of the mid-band resonances in the early part of the decay. At frequencies approaching 500 Hz, it can be seen that the output has dropped below the lowest measured level, roughly 40 dB down, before 350 ms is reached.



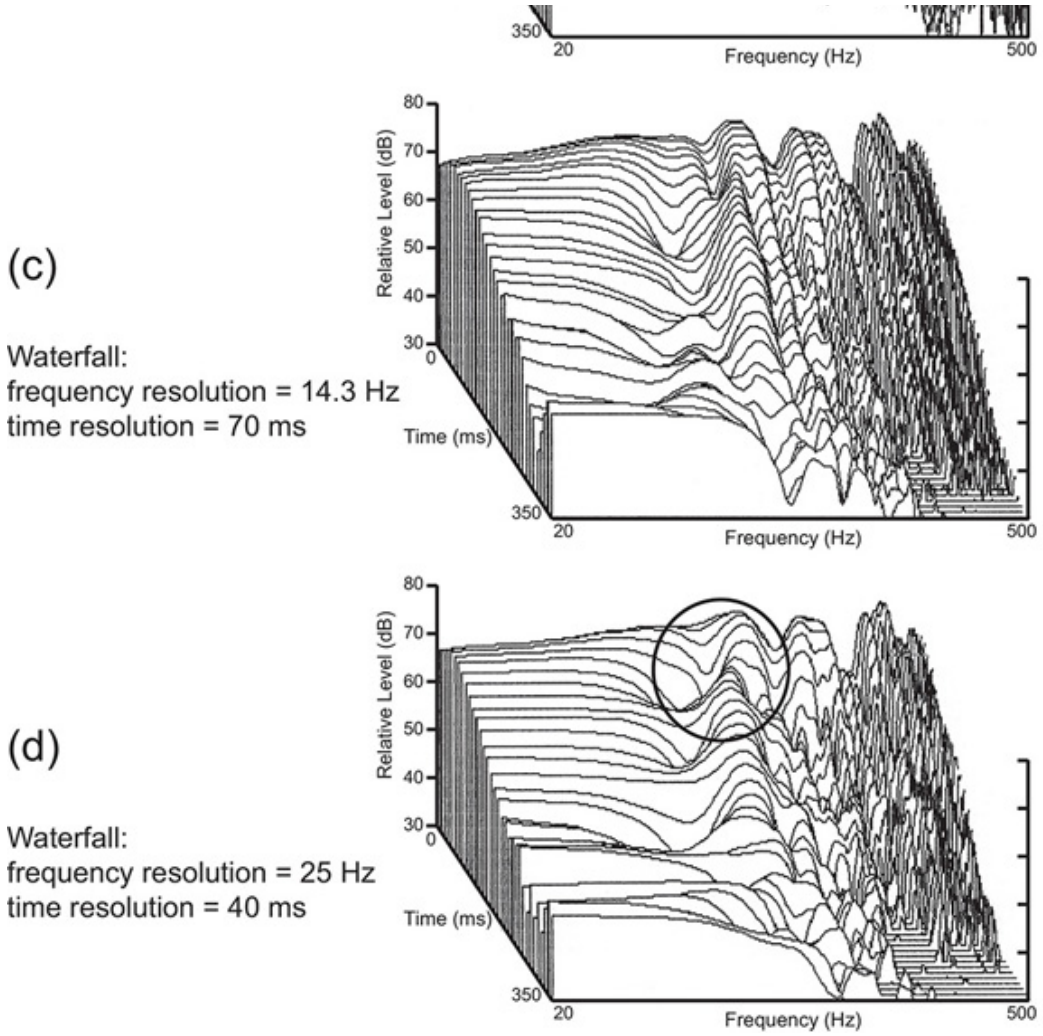


FIGURE 9.11 *Examples of differing resolutions in time and frequency domains, for the same acoustical situation. (a) a very-high-resolution steady-state frequency response. (b), (c) and (d) show waterfall diagrams with different frequency and time resolutions.*

In (d) the top frequency response, and all that follow it, are much smoothed by the further reduced frequency resolution, now 25 Hz. However, because the time resolution is a much-improved 40 ms, it is possible to see more of what is happening in the time domain. Notice in the circled area that what looked like a single decaying resonance in (a) is actually two adjacent resonances with the higher amplitude one decaying very quickly, leaving a higher Q but much lower amplitude one to ring

longer. All the other resonances that looked so alarming in (a) can now be seen to decay into randomness very quickly. In fact, the room appears to have only one resonance with prolonged decay, and its amplitude is more than 10 dB down at the outset—not likely to be a dominant audible factor. This same conclusion can be drawn from the relatively well-behaved steady-state room curve in (a). Good news.

The system looks as though it should have good bass performance, and it did—it was my first home theater where, in 1988, I employed multiple subwoofers to manipulate room resonances. The meaningful measurement is the steady-state response. It reveals the existence of audible resonances that are confirmed in waterfalls, but waterfalls with high-frequency resolution have such poor time resolution that they provide misleading visual information. In audio forum discussions and manufacturers' data there have been numerous examples of erroneous conclusions being drawn from waterfall data of this kind. Usually, the authors don't reveal the parameters of the measurement; in some cases one suspects that they did not know, or know that it mattered. In terms of the audibility of resonances in rooms, the reliable indicator is the steady-state frequency response, not the time-domain response, certainly not a waterfall that artificially prolongs the apparent ringing.

Other mathematical methods exist that reveal different aspects of the time/frequency trade-off (e.g. Wavelets, Wigner-Ville and Gabor distributions) but each requires skilled interpretation, and none that I am aware of has any supportive psychoacoustical data.

9.6.3 Critical Bands, ERB_{N}s and the "Resolution" of the Hearing System

A longstanding belief in the audio industry has been that measurements employing 1/3-octave bands or smoothing are sufficient to reveal the important perceived clues in frequency response curves. In technical terms such measurements are quite low resolution; spinorama measurements have 1/20-octave resolution to reveal potentially audible resonances. What is the truth?

One of the traditional justifications for 1/3-octave spectral analysis has been that this is a rough approximation of perceptual "critical bands" over much of the mid- and high-frequency range. Some people even have argued that we "hear" in critical bands, that this is the "resolution" of the hearing system. Such statements are simplistic and misleading.

The critical band is not the *resolution* of the hearing system, but it is related to the *discrimination* of tones and how multiple tonal components interact with each other, thereby entering the domain of timbre. Roederer (1995) discusses this in detail summarizing that "it plays an important role in the perception of tone quality and provides the basis for a theory of consonance and dissonance of musical intervals."

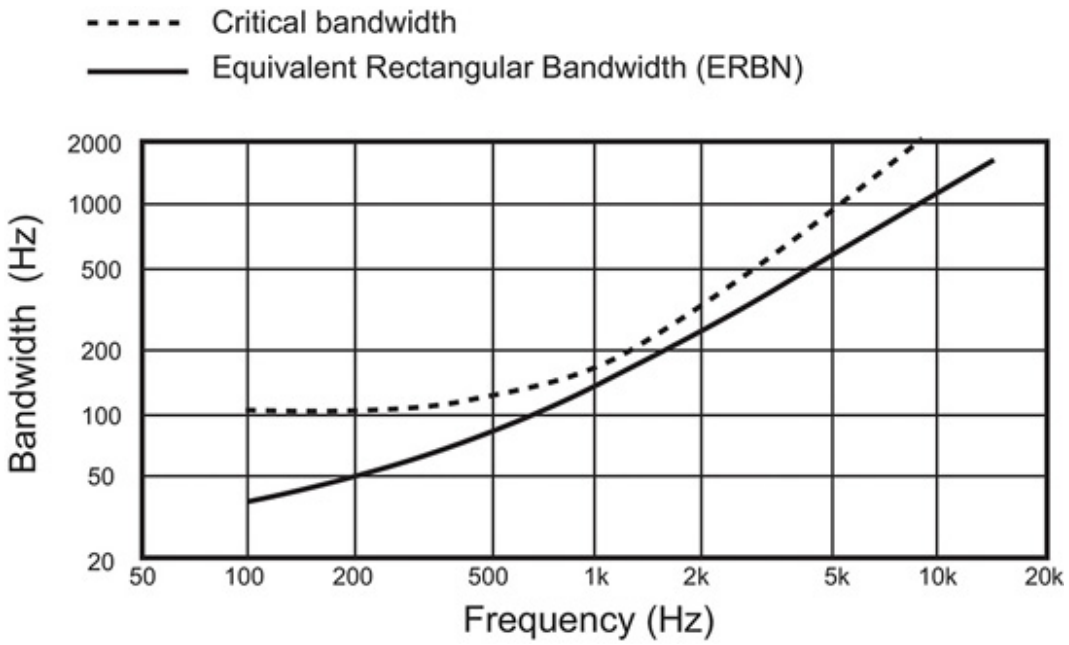


FIGURE 9.12 *The critical bandwidth and ERBN shown as a function of frequency.* ⏎

Readers wishing to understand details of the process are referred to Roederer (1995) and Moore (2003), where the latter leads us to a new definition, the equivalent rectangular bandwidth, the ERBN (see Figure 9.12).

These bands define the bandwidth over which spectral information is summed for estimates of loudness and in the simultaneous masking of tonal signals by broadband noise. But they also define the separation required for two adjacent tones to be individually identifiable. The bandwidth is quite large, ranging from about 40 to 100 Hz at 100 Hz, to about 140 to 180 Hz at 1 kHz, to about 1000 to 2500 Hz at 10 kHz, in a slightly curved line. Within these bands multiple tones (which can be fundamentals and overtones of musical sounds) beat with each other and also generate a perceptual quality called "roughness" (see Figure 9.13). Beats and differences in roughness contribute to the distinctiveness of sounds—their timbre—and so timbre can be changed if the reproducing system has spectral variations occurring within a single critical band or ERBN. The bandwidths of these bands seen in Figure 9.12 indicate that this is easily possible.

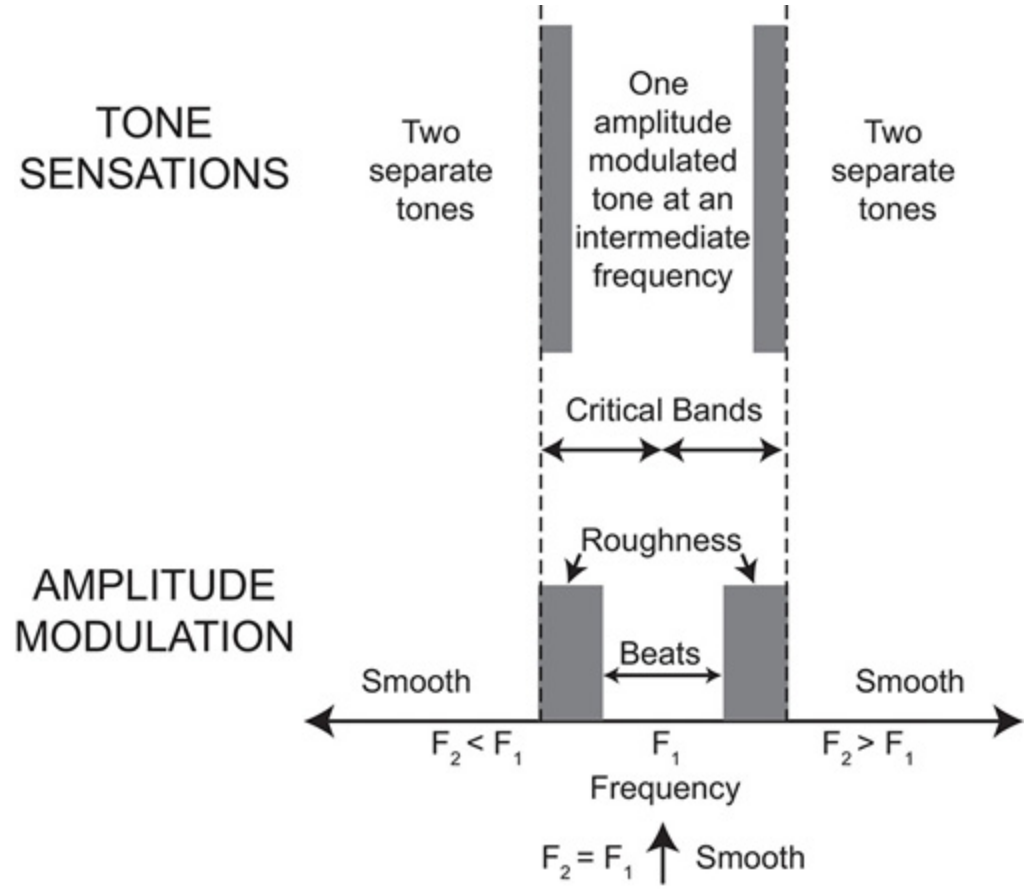


FIGURE 9.13 *An explanation of the pitch and timbre-modifying events that occur within and around critical bands. Inspired by*

figure 2.12 in *Roederer, 1995*. ⏎

A full understanding of the performance of loudspeakers requires measurements with resolution higher than a critical band or ERBN. Most of the anechoic measurements in this book are 1/20-octave, because that was technically convenient, not that it has a unique qualification. There are situations where spectrally smoothed measurements are useful for portraying general trends in spectral balance, as in room curves, but for analytical data collection contemporary understanding suggests that 1/3-octave bands are not optimal (Moore and Tan (2004). The availability of inexpensive or free measurement software that can run on portable devices means that these requirements are not burdensome.

Moore (personal communication, 2008) summarizes: "The auditory filter bandwidths (ERBN values) are about 1/6- to 1/4-octave at medium to high frequencies. And … within-band irregularities in response can have perceptual effects." As has been found in other aspects of perception, it is difficult to find a "one-curve" description for what we hear. These observations emphasize the importance of having high-resolution measurement data that may reveal fluctuations that should ideally be smoothed.

Finally, it almost goes without saying that all of these perceptual factors are changed in hearing-impaired listeners. With hearing loss, the critical bands get wider, and, as a result, even our perception of music is altered. It is not just a matter of hearing or not hearing; it affects how we perceive the sounds that we are able to hear.

9.6.4 The Audibility of Phase Shift and Group Delay

A very long time ago, as part of my self-education about the hearing system, I learned that the basic transduction process in the inner ear performed as a half-wave rectifier. This alone gives one reason to think that acoustic compressions will be heard as different from acoustic rarefactions. We should be sensitive to details in the acoustic waveform as it generates movement of the eardrum and thus initiates the hearing process. Based on

this simple logic, phase shifts and absolute polarity should be audible phenomena. We should be able to hear changes in waveforms.

Naturally I did some tests, reversing polarities of loudspeakers, and introducing phase shift to distort musical waveforms—listening for big differences. They weren't there. At least not in the music I was listening to, through the loudspeakers I was using, from the musical sources I employed. Maybe I was simply unable to hear these things. There were times when I thought I heard subtle differences, but neither one seemed more correct or preferable. Changing loudspeakers made huge differences. Changing records made huge differences. But, the anticipated "dramatic" event of inverting polarity appeared to be missing, despite how appealing the idea of waveform integrity is from an engineering perspective. Are they audible or not, and if so, does it matter in real-world listening?

The combination of amplitude vs. frequency (frequency response) and phase vs. frequency (phase response) totally define the linear (amplitude independent) behavior of loudspeakers. The Fourier transform allows this information to be converted into the impulse response and, of course, the reverse can be done. So, there are two equivalent representations of the linear behavior of systems, one in the frequency domain (amplitude and phase) and the other in the time domain (impulse response).

An enormous amount of evidence indicates that listeners are attracted to linear (flat and smooth) amplitude vs. frequency characteristics; more to be shown later. Figure 7 in Toole (1986) and the excerpts shown here in Figure 2.4 indicate that listeners showed a clear preference for loudspeakers with smooth and flat frequency responses. Figure 2.4 also shows phase responses for those same loudspeakers. It is difficult to see any reliable relationship to listener preference, except that those with the highest ratings had the smoothest curves, but linearity did not appear to be a factor. Listeners were attracted to loudspeakers with minimal evidence of resonances because resonances show themselves as bumps in frequency response curves and rapid up-down deviations in phase response curves. The most desirable frequency responses were approximations to horizontal straight lines. The corresponding phase responses had no special shape, other than the

smoothness. This suggests that we like flat amplitude spectra, we don't like resonances, but we seem to be insensitive to general phase shift, meaning that waveform fidelity is not a requirement.

If one chooses to design a loudspeaker system having linear phase, there will be only a very limited range of positions in space over which it will apply. This constraint can be accommodated for the direct sound from a loudspeaker, but even a single reflection destroys the relationship. One of the loudspeakers in the 7.0 to 7.4 Fidelity Rating category in Figure 2.4 had very linear phase response, but it was not enough to push it into the top category, because it had other problems primarily related to very uneven off-axis frequency response. Therefore, it seems that:

1. Because of reflections in the recording environment there is little possibility of phase integrity in the recorded signal.
2. There are challenges in designing loudspeakers that can deliver a signal with phase integrity over a large angular range, especially vertically.
3. There is no hope of it reaching a listener in a normally reflective room.

All is not lost, though, because two ears and a brain seem not to care.

Many investigators over many years attempted to determine whether phase shift mattered to sound quality (e.g. Hansen and Madsen, 1974a, 1974b; Lipshitz et al., 1982; Van Keulen, 1991; Greenfield and Hawksford, 1990). In every case it has been shown that, if it is audible, it is a subtle effect, most easily heard through headphones or in an anechoic chamber, using carefully chosen or contrived signals. There is quite general agreement that with music, reproduced through loudspeakers in normally reflective rooms phase shift is substantially or completely inaudible. When it has been audible as a difference, when it is switched in and out, it is not clear that listeners had a preference.

Others looked at the audibility of group delay: Blauert and Laws (1978), Deer et al. (1985), Bilsen and Kievits (1989), Krauss (1990), Flanagan et al.

(2005), Møller et al. (2007) finding that the detection threshold is in the range 1.6 to 2 ms, and more in reflective spaces. These numbers are not exceeded by normal domestic and monitor loudspeakers.

Lipshitz et al. (1982) conclude: "All of the effects described can reasonably be classified as subtle. We are *not*, in our present state of knowledge, advocating that phase linear transducers are a requirement for high-quality sound reproduction." Greenfield and Hawksford (1990) observe that phase effects in rooms are "very subtle effects indeed" and seem mostly to be spatial rather than timbral. As to whether phase corrections are needed, without a phase correct recording process, any listener opinions are of personal preference, not the recognition of "accurate" reproduction.

9.6.5 Phase Shift at Low Frequencies—A Special Case

In the recording and reproduction of bass frequencies there is an accumulation of phase shift at low frequencies that arises whenever a high-pass filter characteristic is inserted into the signal path. It happens at the very first step, in the microphone, then in various electronic devices that are used to attenuate unwanted rumbles in the recording environments. More is added in the mixing process, storage systems and in playback devices that simply don't respond to DC. All are in some way high-pass filtered. One of the most potent phase shifters is the analog tape recorder. Finally, at the end of all this is the loudspeaker which cannot respond to DC and must be limited in its downward frequency extension. I don't know if anyone has added up all of the possible contributions, but it must be enormous. Obviously, what we hear at low frequencies is unrecognizably corrupted by phase shift. The question of the moment is, how much of this is contributed by the woofer/subwoofer, is it audible and, if so, can anything practical be done about it? Oh yes, and if so, can we hear it through a room?

Fincham (1985) reported that the contribution of the loudspeaker alone could be heard with specially recorded music and a contrived signal, but that it was "quite subtle". The author heard this demonstration and can concur. Craven and Gerzon (1992) stated that the phase distortion caused by

the high-pass response is audible, even if the cut-off frequency is reduced to 5 Hz. They say it causes the bass to lack “tightness” and become “woolly”. Phase equalization of the bass, they say, subjectively extends the effective bass response by the order of half an octave. Howard (2006) discusses this work, and the abandoned product that was to come from it. There was disagreement about how audible the effect was. Howard describes some work of his own, measurements and a casual listening test. With a custom recording of a bass guitar, having minimal inherent phase shift, he felt that there was a useful difference when the loudspeaker phase shift was compensated for. None of these exercises reported controlled, double-blind listening tests, which would have delivered a statistical perspective on what might or might not be audible, and whether a preference for one condition or the other was indicated.

The upshot of all of this is that, even when the program material might allow for an effect to be heard, there are differences of opinion. It all assumes that the program material is pristine, which it is not, nor is it likely to be in the foreseeable future. It also assumes that the listening room is a neutral factor, which it cannot be. However, if it can be arranged that these numerous other factors can be brought under control, the technology exists to solve the residual loudspeaker issue.

9.6.6 The Audibility of Absolute Polarity—Which Way Is “Up”?

Finally, there is the matter of simple polarity, which is where this discussion started, with the apparent fact that the ear responds differently to pressure compressions and rarefactions. Johnsen (1991), a strong advocate of the importance of polarity, offers an extensive survey of literature, much of it anecdotal, and conducts experiments that are claimed to be definitive evidence of its audibility. Unfortunately, from my perspective, there was a serious flaw in the choice of source material for the subjective tests. Johnsen said: “Sources were LP records exclusively, as the CD medium was found (in collateral, unreported tests) to reveal polarity rather less well.” I have an intimate understanding of the LP medium, and its lengthy

list of non-linear behaviors. It is easy to understand that the asymmetrical distortions added by the playback medium would indeed make polarity changes more easily audible than the demonstrably more linear CD. Were the listeners responding to the medium or the message? It was a flawed experiment.

In contrast, Greiner and Melton (1994), who used CDs as sources, took precautions to avoid contamination of the signals by both source and delivery devices. They employed both synthesized asymmetrical waveforms and naturally asymmetrical sounds. Their conclusion was:

> It is certain from our listening tests that inversion of acoustic polarity is clearly audible for some instruments played in some styles and for some listening conditions. It is not likely that the effects observed in this work were an artifact of the record reproduce system because of the considerable care taken to maintain waveform integrity.
>
> While polarity inversion is not easily heard with normal complex musical program material, as our large-scale listening tests showed, it is audible in many select and simplified musical settings. Thus it would seem sensible to keep track of polarity and to play the signal back with the correct polarity to ensure the most accurate reproduction of the original acoustic waveform.

Greiner went on to say in a subsequent “letters to the editor” exchange that “neither the quality of the transducers nor the loudness level, again within reason, has in my experience a great effect on the audibility of acoustic polarity inversion.”

So, as speculated at the outset, the asymmetrical detection process in the ear can yield audible polarity effects if the signal and the listening conditions are right. The problem is that there appear to be no audio or film industry standards that ensure the delivery of such absolute polarity sounds from the microphones, through the extensive electronic manipulations in control rooms or dubbing stages, which may be different for different components of a mix, and finally through the playback electronics and loudspeakers at home. If there is a polarity switch in the playback chain, one may occasionally find a preferred setting, but there can be no knowing

which setting is correct, and it may be different for different voices and instruments in a band.

9.7 NON-LINEAR DISTORTION

The memory of distortions in 78 rpm recordings and early LPs is still vivid. They ranged from inaudible to intolerable, and many recordings had them. LPs and their playback devices improved, and at their best they became highly enjoyable—except for pesky inner-groove crescendos, the exciting climax of a symphony or opera that one has carefully been prepared for by the composer.

I recall testing phonograph cartridges as part of the effort to improve playback quality (Toole, 1972). This exercise is substantially a test of test records, which really is a test of the entire LP mastering, pressing and playback process. I participated in the creation of a test record, creating some of the sounds and signals recorded on the master tape. Comparisons of the master tape to the LP playback revealed that it was impossible to replay from an LP the same signal that was delivered to the (carefully chosen) mastering lab. When using pure tones or bands of noise, distortions were easily measured and easily audible. They registered in whole (sometimes high) percentages much of the time, and this applied to both harmonic and intermodulation versions. Yet, when the signal was music, the experience was mostly enjoyable. What is going on here?

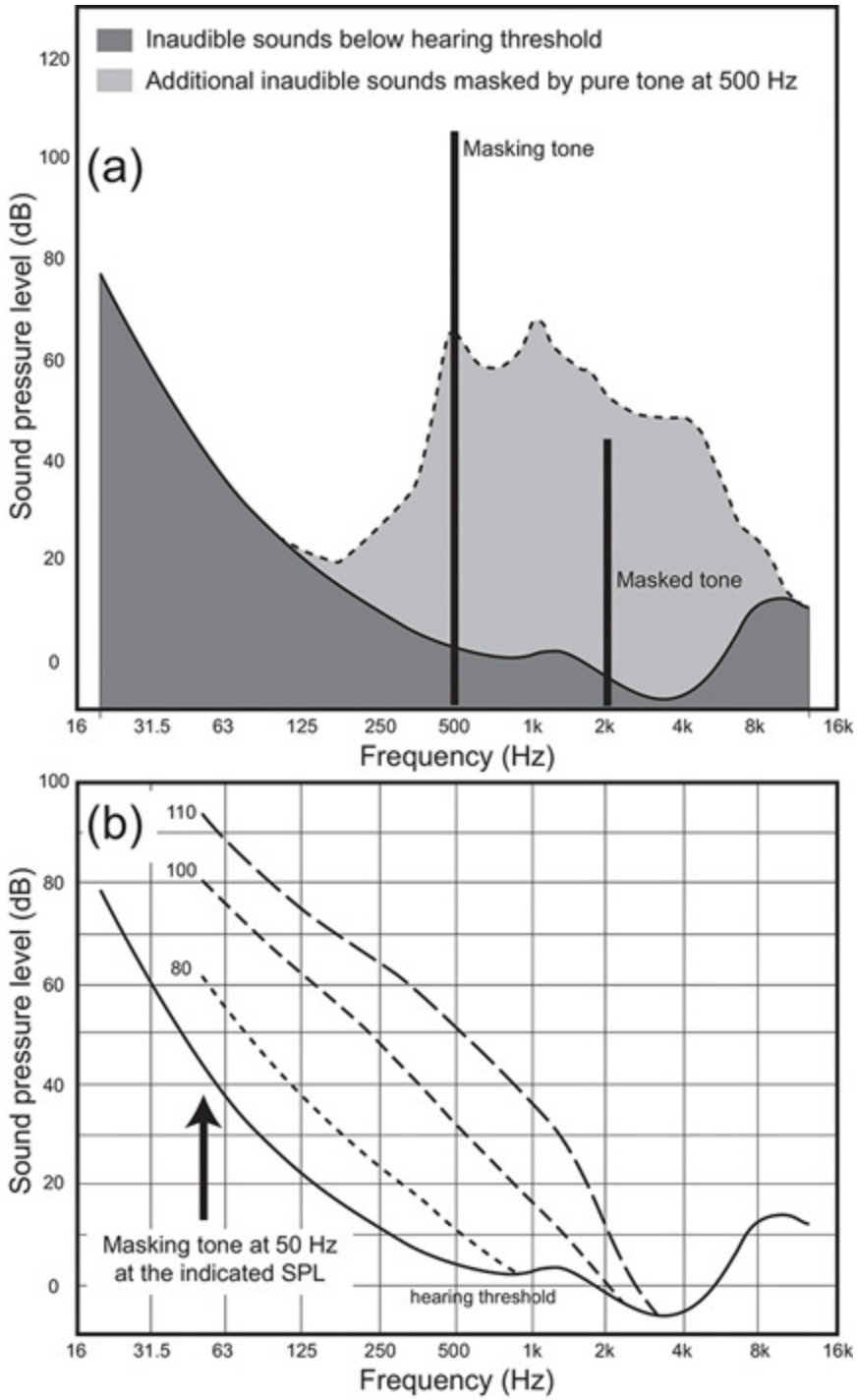


FIGURE 9.14 *(a) A simple view of simultaneous masking of a tone at 2 kHz by another, louder one, at 500 Hz. It shows that the masking effect spreads substantially upward in frequency and only slightly downward. The detailed shape of the dashed masking curve is highly variable, depending on masker level and*

differing among studies but the general effect is as shown. At lower sound levels, the masking effect exhibits less spreading. (b) shows the substantial upward masking effects of very low frequencies, especially at high sound levels. From Fielder, chapter 1.5, figure 1.60 in Talbot-Smith (1999). Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc. ⏎

Masking, of a smaller sound (the distortion) by a larger sound (the same musical signal that created the distortion), as shown in Figure 9.14. Simple test signals leave the distortion products spectrally exposed so that they can be measured, and sometimes enough of them are unmasked that we can hear them. The wide-bandwidth, dense spectrum of music is a much more effective masking signal, in spite of the fact that it is at the same time a generator of much more complex distortion products. It is also almost useless as a measurement test signal.

The masking contours shown in Figure 9.14(b) are especially revealing in that a loud bass sound can have an effect on sounds covering much of the audible spectrum, including the critical voice-frequency range. This may explain why it is sometimes difficult to understand the lyrics in a loud, bass-heavy rock concert, or the dialog in a movie when loud bass rumbles are used for dramatic effect.

To understand what is going on, first consider what the real problem is: non-linear behavior in the audio device. This means that the relationship between the input signal and the corresponding output from the device changes with level. If the input increases by a certain percentage, the output should change by the same percentage, no more, no less. However, the percentage does change, the shapes of audio waveforms are altered from a little to a lot. When we hear these distorted waveforms we notice that new sounds (distortion products) have been created. When the non-linear device is driven by a known, well-specified signal, we can measure how different the output is from the input, and obtain a measure of the magnitude of the distortion—for that specific signal. But, instead of measuring the nature of

the non-linearity itself—as an input-to-output transfer relationship—we have chosen to try to quantify what it does to audio signals. Put a known signal into the device and spectrally analyze what comes out.

Engineers have invented several ways to quantify the magnitude of the non-linearity by measuring the distortion products and expressing them as a percentage of the original input signal, or of that signal plus distortion and noise. If the input signal is a pure tone, the distortion products are harmonic overtones of that signal: harmonic distortion. The problem with that method is that as the fundamental frequency gets higher, the higher harmonics move beyond the hearing range so it ceases to be a relevant metric. Combining two closely spaced tones and scanning them to high frequencies is more useful. Each of the tones will generate a set of harmonic distortions, but they will also interact with each other in the non-linearity and create combination tones—intermodulation distortions. These are sum and difference multiples of the two input signals and their harmonics that extend upwards and downwards in frequency. Because masking is much more effective in the upward frequency direction (see Figure 9.14), it is the difference frequencies that end up being more audible and, because they are inherently unmusical, more annoying.

Thus began the legend that harmonic distortions are relatively benign and intermodulation distortions are bad. It is true in the context of these test signals, but they are both simply different ways of quantifying the *same* problem (the non-linearity in the audio component), and neither test signal (one tone or two) is even a crude approximation of human voices or music. Contributing to the mismatch between perception and measurement is the fact that such a technical measurement totally ignores masking. Included in the numbers generated by the measurements are distortion components that are partially or completely masked. Some of what is measured is inaudible. The percentage distortion numbers cannot accurately anticipate the audibility of the distortion products.

The end result of this is that traditional measures of harmonic or intermodulation distortion are flawed. They do not quantify distortion in a way that can, with any reliability, predict a human response to it while

listening to music or movies. They do not correlate because they ignore any characteristics of the human listener, itself an outrageously non-linear device. The excessive simplicity of the signals also remains a problem. Music and movies offer an infinite variety of input signals and therefore an infinite variety of distorted outputs. The only meaningful target for conventional distortion metrics is "zero". Above that somebody, sometime, listening to something, may be aware of distortion, but we cannot define it in advance.

Taking advantage of advances in psychoacoustic understanding, and utilizing the analytical and modeling capabilities of computers, some investigations are attempting to identify some of the underlying perceptual mechanisms, and develop better test methods. Cabot (1984) provides a good historical perspective; Voishvillo (2006, 2007) provides an excellent overview of the past, present and possible future of distortion measurements. Geddes and Lee (2003), Lee and Geddes (2003, 2006), Moore et al. (2004a), and Temme and Dennis (2016) provide additional useful perspectives.

In loudspeakers it is fortunate that distortion is something that normally does not become obvious until devices are driven close to or into some limiting condition. In large-venue professional devices this is a situation that can occur frequently. In the general population of consumer loudspeakers, it has been very rare for distortion to be identified as a factor in the overall subjective ratings. This is not because distortion is not there, or not measurable, but it is low enough that it is not an obvious factor in judgments of sound quality at normal foreground listening levels—it is masked.

9.8 DYNAMIC RANGE AND POWER COMPRESSION

Today, in much of the music, movies and television we are exposed to, dynamic range is diminishing. Some of it is that programs are tailored for where we listen—cars, earphones on the street, in buses, the subway, etc. where background noises abound.

Occasionally we need to sit down in quiet surroundings, put on an old-fashioned (relatively) uncompressed music source, and be reminded what dynamic range sounds like. It is not "all loud all the time", which some programming seems to be. Sadly, it is part of the "dumbing down" of audio that has been encouraged by streaming services attempting to appeal to the masses. Compressed programs sound tolerable when played through small audio systems that are incapable of playing loud, and are more comprehensible when listening in noisy backgrounds, as in working or traveling. It is a pity, because when heard through capable sound systems they lack the vitality of uncompressed versions—the music survives but the performance does not.

Well recorded music and movies have dynamic range. Dynamic range has two extremes. The low-level limits are influenced by the background noises in the recording and playback venue, as discussed in Section 9.5.1 and illustrated in Figure 9.8. Other than achieving an acoustically quiet room and an adequate signal-to-noise ratio in electronics, there are no challenges in the playback process.

The upper levels of the dynamic range are much more demanding. The right side of Figure 9.3 shows that, depending on the sensitivity of the loudspeakers, amplifier power requirements can be substantial, especially when loudspeaker impedances are low, often less than advertised. This common situation demands high currents from power amplifiers, and not all can provide them without distortion or clipping. As noted in the discussion in Section 9.4, we tend to be quite tolerant of amplifier clipping of peaks, if it is clean (without ringing or artifacts), but that is not always the case. So, the first step is making sure that the chosen power amplifier can drive the chosen loudspeakers to the chosen sound levels at the chosen listening distance. This is not simple.

Obviously, the loudspeakers should be able to radiate the necessary sound power into the listening room without audible distortion or power compression. As has been stated, audible distortion in well-designed loudspeakers is not normally an audible problem until they are driven to their limits. If that happens, distortion products are generated as the

diaphragms are unable to move through the required excursions in a linear manner. Woofers are a common limiting factor as they reach their excursion limits at very low frequencies. When this happens, it affects all frequencies radiated by the woofer, and the effects can be quite audible and unpleasant, especially in smaller systems where the transducer is a woofer/midrange, performing all duties up to the tweeter crossover. This is where bass management and a subwoofer are a huge benefit, allowing the system to play louder, increasing its usable dynamic range.

Tweeters are another weak spot because they sometimes are used to lower than prudent frequencies, and at high sound levels the diaphragm displacements become non-linear, distortion products are generated, and the sound quality degrades. Power compression is a secondary effect that is sometimes noticed.

Power compression is a phenomenon that is related to high sound levels, but is not a non-linear distortion—it is a slowly varying linear distortion. As the voice coils of transducers heat up, their resistance increases and sound level output drops. It is not a very large amplitude change, but it can affect large portions of the frequency range. As the average power output increases, one can measure changes in the frequency response, often sagging output in the tweeter frequency range. Interestingly. How power compression is measured affects how much is revealed. Pure tone measurements at different input power levels exercise the transducers in a manner that is different from program material, and it can matter whether the frequency sweep begins at high or low frequencies. Ideally, one would condition the system with a broadband signal with an energy distribution representing a typical program and, once it has stabilized, interrupt the signal to perform a quick measurement.

Woofers are heavy-duty devices that can be constructed to withstand enormous amounts of power input and physical abuse. Tweeters are inherently more fragile as moving mass needs to be small. For this reason, loudspeakers required to perform at high sound levels for long periods tend to employ multiple conventional transducers or horns. There are options.

In summary, many conventional loudspeakers can sound absolutely superb at moderate sound levels, but, if pushed to excessive sound levels they might begin to sound “loud”. Other designs, engineered to reproduce high sound levels can cause hearing damage without one being aware. The engineering challenge has been to find a balance between sound quality accuracy and sound power capability. Some traditional high-power systems sacrificed the ultimate timbral accuracy, in achieving the necessary sound levels and audience coverage. However, technology advances and there are now loudspeakers that perform well in both categories, an example of which is shown in Figure 9.15.

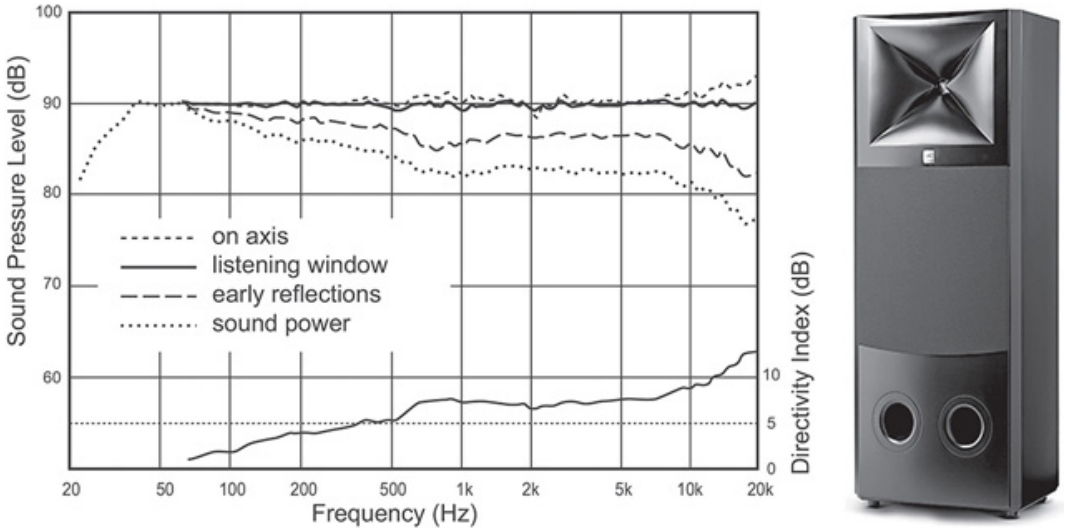


FIGURE 9.15 *The JBL Professional M2 Master Reference Monitor. This is an active loudspeaker, with a dedicated outboard power amplifier/equalizer. The 15-inch woofer is crossed over to a 120° horizontal by 100° vertical horn of unconventional shallow design. The crossover frequency is 800 Hz. It is rated at 117 dB continuous, 123 dB peak at 1 m, 108 dB peak at 8 m distance.* ⏎

The loudspeaker shown in Figure 9.15 is an example of competent engineering. It is a high-power, high–sound quality loudspeaker intended for control-room monitoring, high-end home theaters, and the like. It is

expensive, large, heavy, somewhat "industrial" looking, but it can play louder than humans can safely enjoy and compete with highly rated neutral cone and dome loudspeakers in terms of sound quality.

9.9 WAVELENGTH: THE KEY TO UNDERSTANDING MUCH IN AUDIO

In audio most discussions involve frequency, and events in the frequency domain. But to understand why many of these things happen, it is essential to understand the concept of wavelength. The explanation begins with the speed of sound which, at sea level and normal room temperature (72°F/22°C), is:

- 1131 ft/s (771 mph)
- 345 m/s (1240 km/h)

All audio frequencies travel at the same speed, so the distance traveled in one cycle of sound at a given frequency is the speed of sound divided by the frequency.

Wavelength = Speed of Sound, Frequency
Example: 1131 feet per second divided by 1000 Hz
(cycles per second) = 1.131 feet per cycle = wavelength at 1 kHz

This equation gives the wavelengths for a few frequencies as shown in Figure 9.16. The range of dimensions is huge, meaning that many things may be different at different frequencies.

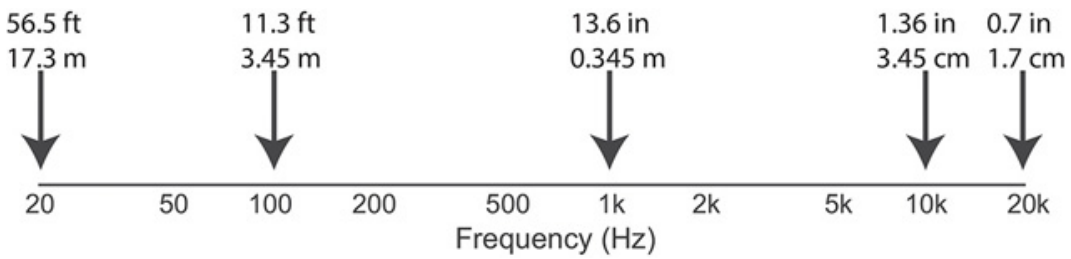


FIGURE 9.16 *Wavelengths at various frequencies.*

9.9.1 The Effects on Loudspeaker Directivity and Sound Quality

Let us consider what this enormous range of wavelengths implies in sound reproduction. When the wavelength of the sound it is radiating is long compared to the dimensions of the sound *pressure* source, the sound will radiate equally in all directions—it is omnidirectional. This is true for all closed-back, enclosed loudspeakers. Dipoles or cardioid pattern loudspeakers behave differently. So, for a 12-inch (0.3 m) subwoofer operating below 80 Hz, the wavelengths are longer than 14.1 ft (4.3 m) and the sound will radiate omnidirectionally. Closed box subwoofers are "pressure" sources. It therefore doesn't matter which way a subwoofer diaphragm faces: the radiated sound will not change.

However, at higher frequencies, shorter wavelengths, the radiated sound will become more focused in the forward direction of the diaphragm, radiating in a progressively narrowing, and possibly irregular, beam with increasing frequency. This is illustrated in a very simplified manner in Figure 9.17(a); when the wavelengths approach and eventually become smaller than the diaphragm of the loudspeaker, it becomes more directional. This is a problem because the off-axis, room reflected, sound cannot have the same spectral balance as the direct sound. This affects sound quality and spatial effects.

A second issue is mechanical wave motion in cone diaphragms themselves, radiating outward from the central voice coil and being reflected back at the periphery. Large diaphragms don't behave well when driven at high frequencies. They acquire "breakup modes"—resonances—that color the sound. Domes have their own resonant modes. Controlling these resonances is a major factor in transducer design. There are two reasons why loudspeakers have woofers, midranges and tweeters, each one optimized to perform well over specific frequency ranges, getting smaller as frequencies rise and wavelengths get shorter. One has to do with the

directional sound radiating characteristics and the other with the timbral neutrality—lack of resonances—in the device.

If the size of the loudspeaker diaphragm is reduced as frequency increases something closer to constant directivity can be achieved, along with much improved frequency and time-domain performance. This is illustrated in Figure 9.17(b). At very low frequencies the bass is radiated omnidirectionally, so the directivity control becomes effective in the low hundreds of Hz. In practice this is a convenience because below the transition frequency what we hear is dominated by room modes, as will be discussed later. The choice of crossover frequency is based on both frequency response and directivity. Ideally, at the crossover frequency the directivity of the low-passed driver and the high-passed driver should be similar. In many modern designs the small tweeters are fitted with shallow waveguides (horns) to assist in acoustical output and to match directivity at crossover.

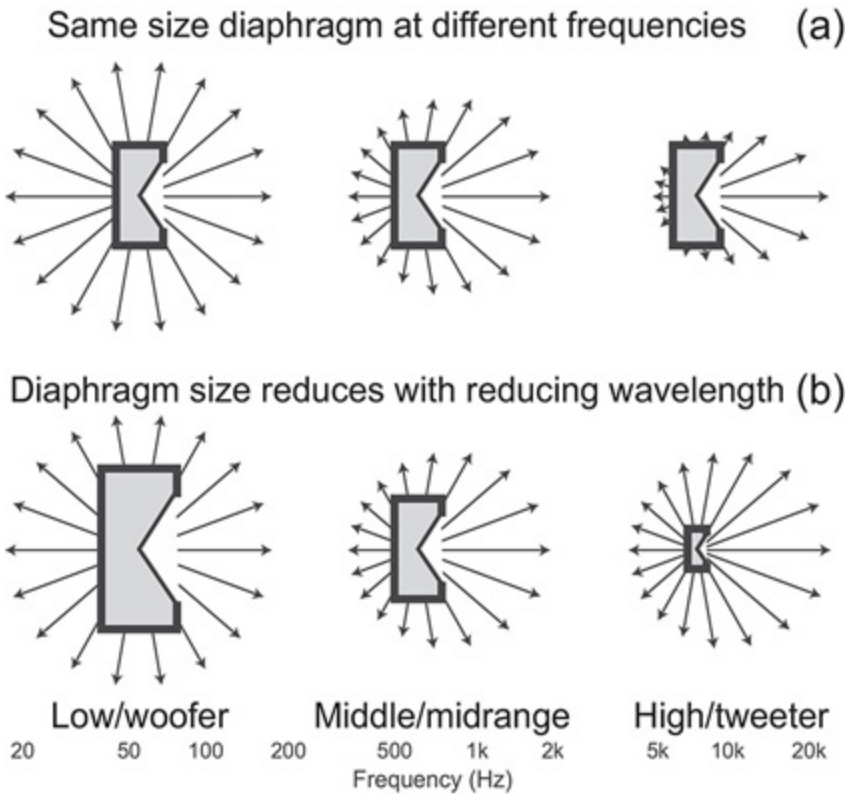


FIGURE 9.17 *A very simplified illustration of sound radiation patterns from loudspeaker drivers of the same size (a) and when*

the sizes are graduated to somewhat correspond to the wavelengths (frequencies) being radiated (b). The goal is relatively constant directivity at frequencies above the transition frequency. Nevertheless, the lowest frequencies will trend to omnidirectionality and the highest frequencies from a conventional tweeter will progressively beam. ⏎

Dividing the frequency range is also advantageous for distortion reduction. Spectrally complex sounds are shared among multiple drivers, thereby reducing the opportunity for intermodulation distortion. More crossovers complicate matters, but improved measurements, waveguides to better match directivities, and computerized crossover design aids have essentially eliminated audible artifacts of crossovers, as can be seen in measurements throughout this book, including Figure 9.15. The advocates of single-driver full-range loudspeaker systems face impossible odds against achieving state-of-the-art performance.

What about electrostatic or electromagnetic large-panel loudspeakers? These loudspeakers are dipole radiators because without an enclosure the opposite polarity radiations from the two sides of the diaphragm cancel each other at the sides, creating nulls and a figure-eight radiation pattern. These appeal to some because it is thought that very low mass diaphragms respond more quickly to musical signals than “heavy” plastic or metal diaphragms. It isn’t true because of the relative motor strengths. The distributed motor systems of panel loudspeakers are not as strong as the linear-motor moving coils of conventional transducers.

More importantly, the film diaphragms have limited movement, so large areas are necessary to generate the required sound pressures. Because directivity is determined by the wavelength relative to the size of the radiating surface, a large surface radiating the full audio bandwidth becomes progressively more directional with increasing frequency. The usual solution for the directivity problem is to subdivide the panel into progressively smaller areas, radiating successively higher frequencies, using crossovers. Many employ a simple narrow tweeter strip. Perhaps the most

ambitious example of this approach was the Quad ESL63 which drove the diaphragm with an arrangement of concentric rings, expanding the source area as frequency decreased, with the aim of generating the wavefront of a point source. Others curve the surface, but the nature of the drive mechanism allows only a gentle curvature.

There is also the belief that, lacking an enclosure, thin film loudspeakers are inherently free from annoying box resonances. That is true, but as Figure 2.14 (loudspeaker M) illustrates, poorly designed thin film diaphragms can also resonate. It was not neutral.

There are practical reasons why cones, domes and horns became and remain the prevalent loudspeaker paradigms.

9.9.2 Acoustical Resonances in Rooms and Enclosures

The following will focus on resonances in rooms, because, for consumers, they are the most serious problem. The underlying principles of resonances in loudspeaker enclosures and rooms are the same but because of smaller dimensions, loudspeaker enclosure resonances are at higher frequencies. They are handled by competent engineering. But customers must solve the room resonance problem.

Sounds of all frequencies and wavelengths reflect around rooms in a somewhat chaotic manner, but when the wavelengths are of the same order as the dimensions of a room, some special events occur. Between parallel surfaces that are half a wavelength apart, or multiples of half-wavelengths apart, the sound traveling in both directions add constructively, reinforcing each other, and a standing wave develops. This is the basis of a room mode or resonance. Figure 9.18 illustrates the basics.

At the frequencies for which the wall separation is a multiple of one-half wavelength, there will be a resonance and a standing wave. These are called axial modes, because they exist along each of the principal axes of a rectangular room: length, width and height.

When there is a standing wave, it is obvious that the sound level at the frequency of the resonance will change as one moves around the room. In fact, a classic demonstration of the phenomenon is to set up a loudspeaker

placed against an end wall radiating a pure tone at the first-order resonance frequency along that dimension of the room, and have listeners walk from one end to the other. In normally reflective rooms, the sound will be about equally loud at both ends of the room and will almost disappear at the halfway point. A variation of this is to have the listeners stand at a single point and scan the signal generator through a range of low frequencies—there will be huge fluctuations in loudness, high at some frequencies, low at others. It is a simple and persuasive demonstration of the problem confronting us.

This topic will be elaborated on in a later chapter, but for now it is sufficient to understand that it is the wavelength of sound relative to the dimensions of rooms that determine whether and at what frequencies there will be resonances and the accompanying standing waves. The resonant modes illustrated here occur between two parallel surfaces and exist along the principal axes of rooms: length, width and height. They are called axial modes, and although other kinds of modes will be shown to exist, these are the dominant factors in most rooms.

It may seem odd that Figure 9.18(d) shows "particle velocity" when what we hear and normally measure is sound pressure. It turns out that the vibratory movement of air molecules is important in understanding how resistive absorbers (fibrous tangles and acoustical foam) work—the next topic.

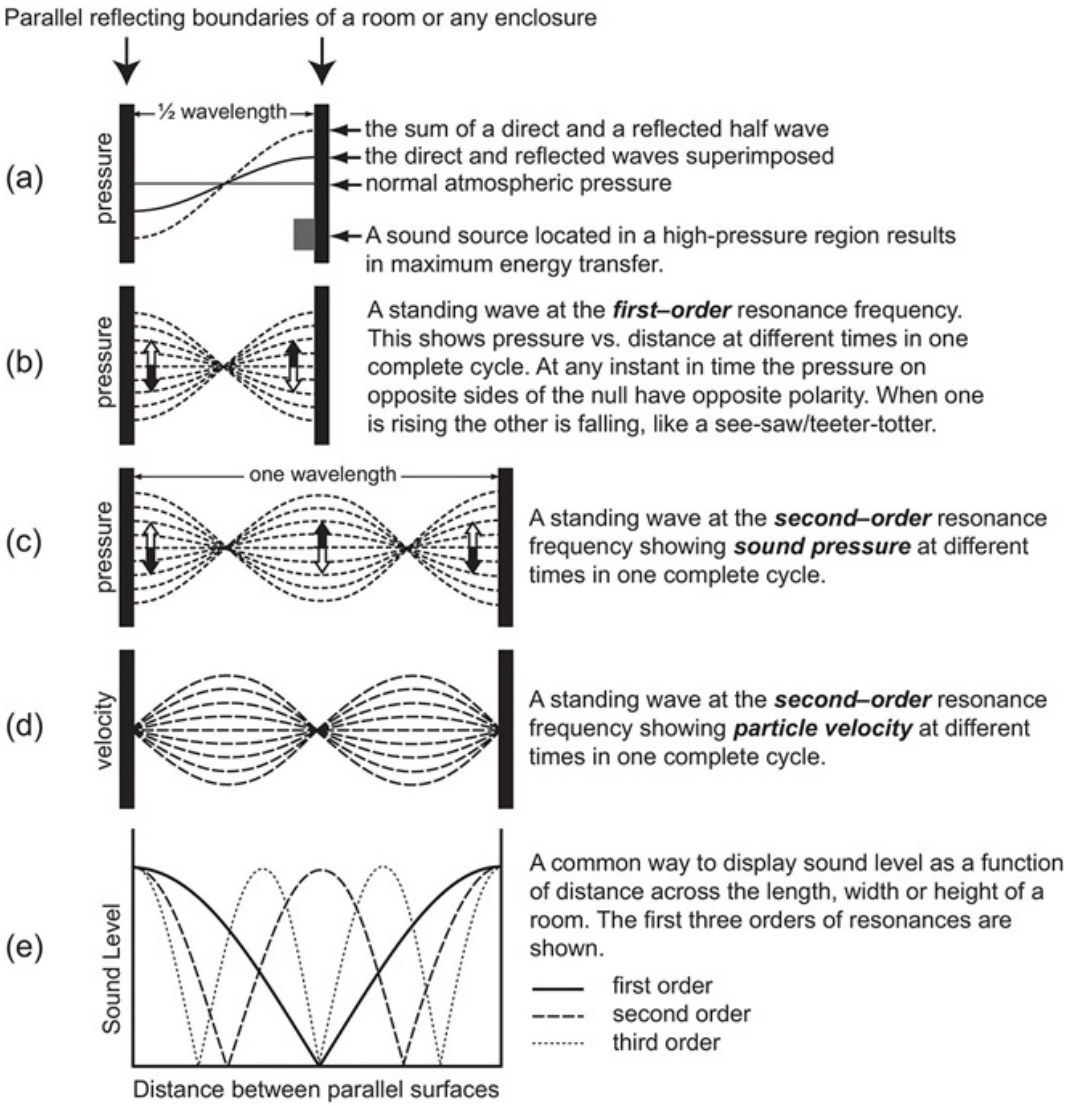


FIGURE 9.18 *(a) The mechanism by which a standing wave is formed, illustrated in a "stop-action" view. At the frequency for which the distance between the walls is exactly one-half wavelength, the direct sound wave traveling towards one wall is precisely replicated by the reflected sound traveling away from it. They add, producing a higher amplitude resultant, which is called a "standing wave" because it has a constant form. In (b) is an attempt to show what happens in a running situation. The "stop-action" resultant waveform shown in (a) cycles up and down with time, once per period of the signal, and it is shown at several points in the cycle. In this first-order standing wave along one*

axis of a room there is one null, one location where it is possible to stand and hear almost nothing—in the center of the room. Moving towards either wall the sound gets louder. It is very important to note that, at any instant in time, at any one of the "stop-action" curves, when the sound pressure on one side of the null is increasing, the sound pressure on the other side is decreasing (shown by the white and black arrows). (c) shows the situation when the distance between the walls is one wavelength. This pattern can also exist between the walls in (a), but at double the frequency. (d) shows the distribution of particle velocity as a function of distance. This is important to understanding resistive absorbers. (e) shows a common manner of representing the sound pressure distribution across a room which is graphically simpler, but it must be remembered that there is a polarity reversal at each null. This will be important in multiple-subwoofer mode-attenuating schemes. ⏎

9.9.3 Understanding Resistive and Membrane Absorbers

Most of the absorbers we use are resistive: messy tangles of glass, cellulose or mineral wool fibers, or leached (open-cell) foam that creates a similarly complex path for sound to travel through. Sound is a pressure wave that propagates through an elastic medium, the compressible fluid air in this case, creating compressions and rarefactions as it passes. The air molecules move back and forth, oscillating in place, in the direction of sound propagation, but not traveling. The wave motion travels at the speed of sound, but the molecules themselves just vibrate, passing their energy on to their neighbors as the sound wave passes. If this movement is forced to happen inside a resistive absorber, some energy is lost in the turbulence of moving within the fibrous tangle—the energy is "absorbed" by being converted to heat. Hence the description of the material: porous, and the mechanism of absorption: resistance. But the resistive losses can happen

only if the air molecules are in motion and the porous material has significant flow resistance.

Looking at Figure 9.18(c) and (d) it is evident that pressure and particle velocity are inverses of each other. It is intuitively easy to understand that air molecules that are against a hard wall have nowhere to go, and particle velocity is zero. Equally instinctive is the notion that a sound pressure wave arriving at a rigid wall will build up pressure as it is unsuccessful at moving the wall and is reflected. So, at a reflecting boundary, pressure is at a maximum, and particle velocity is at a minimum.

Thus, we reach the first level of understanding how best to absorb sound. If resistive absorption is to be useful, it must be at a distance from the room boundary—the portion of it that is adjacent to the boundary is doing nothing because the particles are not moving. The effectiveness increases with distance from the boundary—the thickness of the material—up to a maximum at the 1/4-wavelength distance where the particle velocity is maximum (Figure 9.18(d)). Look at Figure 9.16 and divide those wavelengths by four to get an idea of why resistive absorbers are not attractive solutions at low frequencies—even at a mid-bass frequency of 100 Hz the 1/4-wavelength is 2.8 ft (0.86 m) which seriously intrudes on expensive real estate. This explains why it is the solution of choice at middle and high frequencies. A 1/4-wavelength at 1000 Hz is 8.6 cm (3.4 inches), and at 10 kHz it is 0.86 cm (0.3 inches). A later chapter will show measured absorption coefficients for various materials which will help in selecting absorbers, and how an empty or fluff-filled space behind an absorbing panel can improve performance.

The popular alternative at low frequencies is membrane absorption. When the pressure builds up at the room boundary, if the boundary is flexible it will move, thereby extracting energy from the sound field and reflecting less back into the room. Again, sound energy is turned into heat in the frictional losses of the moving wall surface. When you feel the bass in the floor or walls, membrane absorption is happening. Naturally, if the room boundaries are massive and stiff, as in masonry construction, or in sound-isolated rooms, and not adequately absorbing, membrane absorbers

must be added. Most of these are tuned to absorb sounds over specific bands of frequencies, sometimes narrow bands, sometimes wider bands. To know what to build or buy, it is important to know what frequencies are in need of attenuation so that (a) the right membrane absorber is used and (b) it is possible to find the high-pressure regions for that frequency in the standing wave patterns of the room, and place the absorber at a high-pressure point where it will be most effective. Corners are popular locations as pressure maxima of many modes combine at those locations.

All of this involves the concept of wavelength.

9.9.4 Diffusers and Other Sound-Scattering Devices

Sound-scattering, diffusing surfaces are necessary in performance spaces where there is a fixed amount of sound energy from voices and musical instruments and it is necessary to deliver that sound to all parts of the auditorium or to microphones in a recording studio. Sound reproduction is a very different process.

Do we want a diffuse sound field in our listening rooms? No, it is the purpose of multichannel audio systems to deliver correlated and uncorrelated sounds to our ears, as required to generate perceptions of localized sound sources and ambiguous spatial and enveloping effects. These are decided by recording engineers. In small listening rooms and home theaters, sound-scattering and diffusing surfaces are used as an alternative to absorption in attenuating and redistributing strong reflections, or flutter echoes between parallel surfaces.

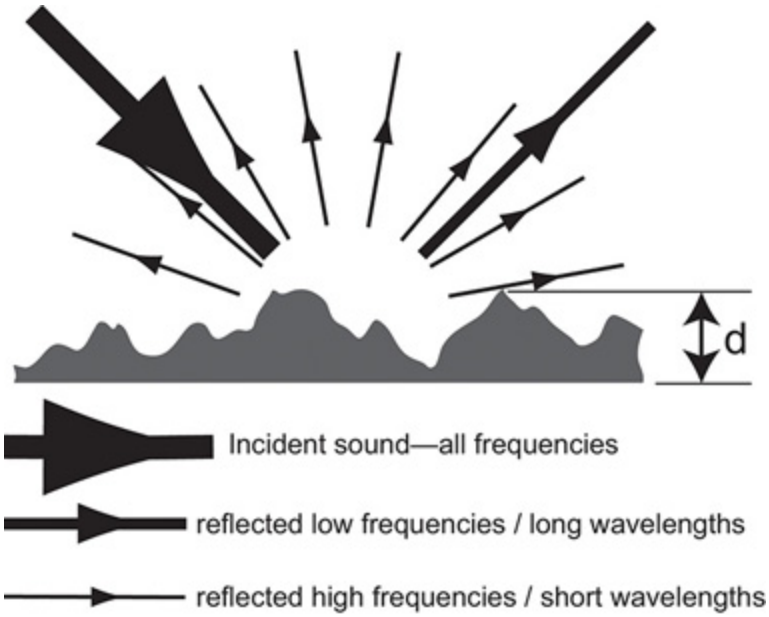


FIGURE 9.19 *A sketch showing the basic concept of a scattering/diffusing surface. At frequencies with wavelengths long compared to "d", the depth of the random protruding features in the irregular surface, there is no diffusion and the sound is simply reflected as if the surface were flat. At short wavelengths, sounds are scattered in many directions simultaneously. The sound sent in any one direction is attenuated, which itself is a reason to use these devices. Engineered versions of this random surface function much more efficiently and can be directional.*

The classic demonstration is the "handclap" test. Stand between two reflective parallel walls, clap your hands and (especially to the person clapping) there may be a disturbing flutter echo: the clap rattling back and forth between the walls. This is an interesting demonstration, but the real problem is when the "clap" emerges from a loudspeaker and is heard as a flutter echo to a person sitting in the audience. So, the first test involves someone moving from speaker to speaker in a system, clapping the hands, while someone else moves around the listening area to see if there is a problem. Then, if there is, find the offending reflecting surface and treat it. Absorption is one solution, scattering or diffusing the energy is an alternative.

Handclaps tend to exhibit a spectral maximum around 2 kHz and higher, so it takes very little to damp them. From the preceding discussion, 1 inch (25 mm) of resistive absorber will do the job. If a diffusing surface is decided on, there is again a decision about the thickness, and again the answer will be based on wavelength.

Figure 9.19 illustrates the basic principle underlying diffusing surfaces. Each diffuser, depending on its design and depth, will have a frequency below which it does not scatter the sound. It is therefore desirable that the scattered/diffused sound should cover the bandwidth above the transition frequency, to avoid distorting the spectrum of the reflected sound field.

A crude guideline is that the depth of a geometrical (e.g. semi-cylinders, multifaceted shapes) diffuser should be not less than about a foot (300 mm), and that of an engineered (e.g. Schroeder, QRD) diffuser should be not less than about 8 inches (200 mm) deep. Gilford (1959) concluded that to be useful, projections from walls must be 1/7-wavelengh or greater at the lowest frequency of interest. At 300 Hz, that is about 7 inches (180 mm). For handclaps (say 1 kHz and up), 2 inches (50 mm) will suffice—but we listen to more than handclaps. Textured paint or plaster are visually interesting but acoustically of no value.

If, for example, the diffuser has a lower limit of 1 kHz, it means that the sound sent off in the direction of the main (specular) reflection arrow will have a lot of energy up to the diffusion limit, and then the high frequencies in the directional reflection will be attenuated. In other directions, listeners will hear the reverse: high frequencies with insufficient spectral energy in the lower frequencies. It all depends on the specific circumstances, to be sure, but one cannot ignore the fact that only a portion of the spectrum is being manipulated by these devices. As the experts, Cox and D'Antonio (2009), caution us there is more to the acoustical performance of these devices than meets the eye. They become quite effective absorbers when covered with fabric, as is often done for visual reasons. When sellers offer only dimensions measurable with a tape measure as "specifications", one needs to be suspicious.

REFERENCES

ANSI/ASA S12.2-2008. "Criteria for Evaluating Room Noise", *Standards Secretariat of the Acoustical Society of America*, Melville, NY. ⏎

Bilsen, F. A. and Kievits, I. (1989). "The Minimum Integration Time of the Auditory System", *86th Convention, Audio Eng. Soc.*, Preprint 2746. ⏎

Blauert, J. and Laws, P. 1978). "Group Delay Distortions in Electroacoustical Systems", *J. Acoust. Soc. Am.*, vol. 63, pp. 1478–1483. ⏎

Broner, N. (2004). "Rating and Assessment of Noise", *Australian Institute of Refrigeration, Air Conditioning and Heating Conference*. www.AIRAH.org.au ⏎

Cabot, R. C. (1984). "Perception of Nonlinear Distortion", *Audio Eng. Soc. 2nd International Conference*, Preprint C1004. ⏎

Cohen, E. and Fielder, L. (1992). "Determining Noise Criteria for Recording Environments", *J. Audio Eng. Soc.*, vol. 40, pp. 384–402. ⏎

Cox, T. and D'Antonio, P. (2009). *Acoustic Absorbers and Diffusers*, 2nd edition, Taylor and Francis, Abingdon, Oxon, England. ⏎

Craven, P. G. and Gerzon, M. A. (1992). "Practical Adaptive Room and Loudspeaker Equalizer for Hi-Fi Use", *92nd Convention, Audio Eng. Soc.*, Preprint 3346. Also: UK DSP Conference, Paper DSP-12. ⏎

Deer, J. A., Bloom, P. J. and Preis, D. (1985). "New Results for Perception of Phase Distortion", *77th Convention, Audio Eng. Soc.*, Preprint 2197. ⏎

Fincham, L. R. (1985). "The Subjective Importance of Uniform Group Delay at Low Frequencies", *J. Audio Eng. Soc.*, vol. 33, pp. 436–439. ⏎

Flanagan, S., Moore, B. C. J. and Stone, M. (2005). "Discrimination of Group Delay in Clicklike Signals Presented via Headphones and Loudspeakers", *J. Audio Eng. Soc.*, vol. 53, pp. 593–611. ⏎

Fletcher, H. and Munson, W. A. (1933). "Loudness, Its Definition, Measurement and Calculation", *J. Acoust. Soc. Am.*, vol. 5, pp. 82–108. ⏎

Geddes, E. R. and Lee, L. W. (2003). "Auditory Perception of Nonlinear Distortion—Theory", *115th Convention, Audio Eng. Soc.*, Preprint 5890.

⏎
Gilford, C. L. S. (1959). “The Acoustic Design of Talks Studios and Listening Rooms”, *Proc. Inst. Electrical Eng.*, vol. 106, pp. 245–258. Reprinted in *J. Audio Eng. Soc.*, vol. 27, pp. 17–31, 1979. ⏎
Greenfield, R. and Hawksford, M. (1990). “The Audibility of Loudspeaker Phase Distortion”, *88th Convention, Audio Eng. Soc.*, Preprint 2927. ⏎
Greiner, R. and Melton, D. (1994). “Observations on the Audibility of Acoustic Polarity”, *J. Audio Eng. Soc.*, vol. 42, pp. 245–253. And, letters to the editor: *JAES*, vol. 43, pp. 147–149. ⏎
Hansen, V. and Madsen, E. R. (1974a). “On Aural Phase Detection”, *J. Audio Eng. Soc.*, vol. 22, pp. 10–14. ⏎
Hansen, V. and Madsen, E. R. (1974b). “On Aural Phase Detection: Part II”, *J. Audio Eng. Soc.*, vol. 22, pp. 783–788. ⏎
Howard, K. (2006). “Wayward Down Deep”, *Stereophile*, vol. 29, July. www.stereophile.com/reference/706deep/index.html ⏎
ISO 226 (2003). © ISO. This material is reproduced from ISO 226 with permission of the American National Standards Institute (ANSI) on behalf of the International Organization for Standardization (ISO). No part of this material may be copied or reproduced in any form, electronic retrieval system or otherwise or made available on the Internet, a public network, by satellite or otherwise without the prior written consent of the ANSI. Copies of this standard may be purchased from ANSI, 25 West 43rd St., New York, NY 10036. http://webstore.ansi.org. ⏎
Johnsen, C. (1991). “Proofs of an Absolute Polarity”, *Audio Eng. Soc. 91st Convention*, Preprint 3169. ⏎
Kommamura, M. and Mori, S. (1983). “New Measurement of Frequency Response Flatness Limen”, *J. Acoust. Soc. Jpn.*, vol. 39, p. 463. ⏎
Krauss, G. J. (1990). “On the Audibility of Group Distortion at Low Frequencies”, *88th Convention, Audio Eng. Soc.*, Preprint 2894. ⏎
Lee, L. W. and Geddes, E. R. (2003). “Auditory Perception of Nonlinear Distortion”, *115th Convention, Audio Eng. Soc.*, Preprint 5891. ⏎
Lee, L. W. and Geddes, E. R. (2006). “Audibility of Linear Distortion with Variations in Sound Pressure Level and Group Delay”, *115th Convention,*

Audio Eng. Soc., Preprint 6888. ⏎

Lipshitz, S., Pocock, M. and Vanderkooy, J. (1982). "On the Audibility of Midrange Phase Distortion in Audio Systems", *J. Audio Eng. Soc.*, vol. 30, pp. 580–595. Comments by Shanefield, D. and Authors' Reply (1983), vol. 31, pp. 447–448. Comments by Moir, J. (1983), vol. 31, p. 939. More comments by van Maanen, H. R. E., Shanefield, D. and Moir, J. (1985), vol. 33, pp. 806–808. ⏎

Mäkivirta, A. V. and Anet, C. (2001). "A Survey Study of In-Situ Stereo And Multi-channel Monitoring Conditions", *111th Convention, Audio Eng. Soc.*, Preprint 5496. ⏎

Møller, H., Minnaar, P., Olesen, S. Christensen, F. and Plogsties, J. (2007). "On the Audibility of All-Pass Phase in Electroacoustical Transfer Functions", *J. Audio Eng. Soc.*, vol. 55, pp. 115–134. ⏎

Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing*, 5th edition, Academic Press, London. ⏎

Moore, B. C. J. and Tan, C-T. (2004). "Development and Validation of a Method for Predicting the Perceived Naturalness of Sounds Subjected to Spectral Distortion", *J. Audio Eng. Soc.*, vol. 52, pp. 900–914. ⏎

Moore, B. C. J., Tan, C-T, Zacharov, N. and Mattila, V-V. (2004a). "Measuring and Predicting the Perceived Quality of Music and Speech Subjected to Combined Linear and Nonlinear Distortion", *J. Audio Eng. Soc.*, vol. 52, pp. 1228–1244. ⏎

Olive, S. E. (1994). "A Method for Training Listeners and Selecting Program Material for Listening Tests", *Audio Eng. Soc. 97th Convention*, Preprint 3893. ⏎

Olive, S. E. (2004a). "A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 1—Listening Test Results", *116th Convention, Audio Eng. Soc.*, Preprint 6113. ⏎

Olive, S. E. (2004b). "A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 2—Development of the Model", *117th Convention, Audio Eng. Soc.*, Preprint 6190. ⏎

Pollack, I. (1952). “The Loudness of Bands of Noise”, *J. Acoust. Soc. Am.*, vol. 24, pp. 533–538. ⏎

Robinson, D. W. and Dadson, R. S. (1956). “A Re-determination of the Equal-loudness Relations for Pure Tones”, *Br. J. Appl. Phys.*, vol. 7, pp. 166–181. ⏎

Roederer, J. G. (1995). *The Physics and Psychophysics of Music*, 3rd edition, Springer-Verlag, New York. ⏎

Stevens, S. S. (1957). “Calculating Loudness”, *Noise Control*, vol. 3, pp. 11–22. ⏎

Stevens, S. S. (1961). “Procedure for Calculating Loudness: Mark VI”, *J. Acoust. Soc. Am.*, vol. 33, pp. 1577–1585. ⏎

Talbot-Smith, M., editor (1999). *Audio Engineer’s Reference Book*, Focal Press, Oxford. ⏎

Temme, S. and Dennis, P. (2016). “In Vehicle Audio System Distortion Audibility versus Level and Its Impact on Perceived Sound Quality”, *141st Convention, Audio Eng. Soc.*, Preprint 9651. ⏎

Tocci, G. C. (2000). “Room Noise Criteria—the State of the Art in the Year 2000”. www.cavtocci.com/portfolio/publications/tocci.pdf. ⏎

Toole, F. E. (1972). “Understanding the Phono Cartridge”, *Electron Magazine (Canada)*, three parts, in: August, September and October. ⏎

Toole, F. E. (1973). “Loudness—Applications and Implications to Audio”, *dB The Sound Engineering Magazine*, pt. 1, May, pp. 27–29, pt. 2, June, pp. 25–28. ⏎

Toole, F. E. (1986). “Loudspeaker Measurements and Their Relationship to Listener Preferences”, *J. Audio Eng. Soc.*, vol. 34, pt. 1, pp. 227–235, pt. 2, pp. 323–348. ⏎

Van Keulen, W. (1991). “Possible Mechanism for Explaining Monaural Phase Effects of Complex Tones”, *90th Convention, Audio Eng. Soc.*, Preprint 3065. ⏎

Voishvillo, A. (2006). “Assessment of Nonlinearity in Transducers and Sound Systems—From THD to Perceptual Models”, *121st Convention, Audio Eng. Soc.*, Preprint 6910. ⏎

Voishvillo, A. (2007). “Measurements and Perception of Nonlinear Distortion—Comparing Numbers and Sound Quality”, *Audio Eng. Soc., 123rd Convention*, Preprint 7174. ⏎

Warnock, A. C. C. (1985). “Introduction to Building Acoustics”, *Canadian Building Digest—CBD-236*, published by National Research Council Canada, Institute for Research in Construction. http://irc.cnrc.gc.ca/cbd236e.html ⏎

CHAPTE
R 10

Subjective Measurements Turning Opinion Into Fact

Floyd Toole

DOI: 10.4324/9781003477495-11

Listening tests are, by definition, subjective. They involve all the remarkable human listening skills, but also the imperfections. The opinions that result may have meaning at the moment, but experience has shown that our opinions can change with time and exposure to information. Reviewers write about their opinions in long, articulate essays dwelling on the details of recordings revealed by the loudspeakers they are listening to. If you get to hear those loudspeakers, their opinions may influence your own, or even influence the decision of what to buy. Some opinions matter more than others, but which ones does one trust?

Those readers who have absorbed previous chapters in this book will appreciate that if one has trustworthy opinions, they can provide useful guidance in identifying useful measurements and interpreting them. Getting trustworthy measurements is a technical and financial challenge, but they are

becoming more available; see Section 2.9. Getting trustworthy subjective data is much more difficult—usually we get opinions. An opinion is a judgment that is not conclusive, as opposed to facts. Turning opinions into facts is the challenge.

My very early double-blind listening tests employed the simplest metric of variation, standard deviation. In those days, using a 10-point scale of "Fidelity", the very best listeners were repeating their ratings over extended periods of time with standard deviations around 5%. I found this remarkable, because at that time analog VOM (volt-ohm-milliammeter) multi-testers had similar specifications. Yes, these had the characteristics of measurements.

Subjective measurements have provided the entry point to understanding the psychoacoustic relationships between what we measure and what we hear. The key to getting useful data from listening tests is in controlling or eliminating all factors that can influence opinions other than the sound itself. This is the element that has been missing from much of the audio industry, contributing to the belief that we cannot measure what we hear. Under the right circumstances, humans can be remarkably good measuring instruments.

It is often assumed that listeners are recognizing excellence and rejecting inferiority when judging sound quality. Regular concertgoers and musicians were assumed to be inherently more skilled as listeners. As logical as this seemed, it was soon thrown into question when listeners in the tests showed that they could rate products just as well with studio-created popular music as they could with classical music that was painstakingly captured with simple microphone setups, frequently even better. How could this be possible? None of us had any idea how the studio creations should sound, with all of the multitrack, close-miked, pan-potted image building and signal processing that went into them. The explanation was in the comments written by the listeners. They commented extravagantly on the problems in the poorer products, heaping scorn rich in adjectives and occasional expletives on things that were not right about the sound. In contrast, high-scoring products received only a few words of simple praise.

And what about musicians? Their performances in listening tests varied, many were not distinctive, but others were. Some had the occupational handicap of hearing loss, others were clearly more interested in the music than

in the sound and could find “valid interpretations” of instrumental sound in even quite poor loudspeakers. They were articulate, had the right vocabulary to describe what they heard, but their opinions did not always align with those of “ordinary” listeners. The best among them often had audio as a hobby (Gabrielsson and Sjögren, 1979; Toole, 1985).

What about professional recording engineers? They create the recordings we listen to and are very experienced. Those with normal hearing perform similarly to other experienced listeners and audiophiles. However, hearing loss is an occupational hazard, as will be discussed.

The ability to separate what the loudspeakers were doing to the program from the program itself is determined by the experimental method, as explained in Section 1.4. This may be worth revisiting. Multiple-comparison experiments are very effective. In these, listeners were able to freely switch the signal among three or four different products while listening to the music. Thus, the “personalities” of the loudspeakers were revealed through the ways the program changed, while the contribution of the listening room was constant, and that of the program was constant during each program selection. In a single-stimulus, take-it-home-and-listen-to-it kind of test, this would not be obvious. Isolated A vs. B comparisons fail to reveal problems that are common to both test sounds. The experimental method can decide what listeners are able to hear. The following is a repetition from Section 1.4:

> All evaluations yield opinions, but not all opinions are answers to the question being asked.

Humans are remarkably observant creatures, and we use all our sensory inputs to remain oriented in a world of ever-changing circumstances. So, when asked how a loudspeaker sounds, it is reasonable that we instinctively grasp for any relevant information to put ourselves in a position of strength. A knowledgeable audio person could look at the loudspeaker, recognize the brand and perhaps even the model, possibly remember hearing it on a previous occasion and the opinion formed at that time, perhaps recall a review, and, of course, would have at least an approximate idea of the cost. Who among us

has the self-control to ignore all of that and to form a new opinion solely based on the sound being heard at that moment? Our opinions are influenced by a lot of subconscious "baggage".

The following sections will attempt to provide insight and guidance in the conduct of useful listening tests. There are important issues of what questions to ask the listeners, scaling the results, statistical analysis, and so on. It has become a science unto itself. (See Toole and Olive, 2001, and Bech and Zacharov, 2006, for elaborations of experimental procedures and statistical analysis of results.) However, as has been found, even rudimentary experimental controls and elementary statistics can take one a long way.

10.1 SEEING IS BELIEVING: AN AUDIO PLACEBO EFFECT?

Knowledge of the products being evaluated is a powerful source of psychological bias. In comparison tests of many kinds, especially in wine tasting and drug testing, considerable effort is expended to ensure the anonymity of the devices or substances being evaluated. If the mind thinks that something is real, the appropriate perceptions or bodily reactions can follow—the classic placebo effect. In audio, many serious people persist in the belief that they can ignore such non-auditory factors as price, size, brand, and so on.

This is especially true in the few "great debate" issues, where it is not so much a question of how large a difference there is but whether there *is* a difference (Clark, 1981, 1991; Lipshitz, 1990; Lipshitz and Vanderkooy, 1981; Nousaine, 1990; Self, 1988). In controlled A/B/X listening tests and in measurements, electronic devices in general, speaker wire, and audio-frequency interconnection cables are found to exhibit small to nonexistent differences. Any audible differences are usually traced to measurable differences. Yet, some reviewers can write pages of descriptive text about audible qualities in detailed, absolute terms, sometimes alluding to causes that defy the laws of physics. The evaluations reported on were usually done without controls because it is believed that disguising the product identity *prevents* listeners from hearing differences. In fact, some audiophile

publications heap scorn on double-blind tests that fail to confirm their opinions—the data must be wrong.

> In science, contrary evidence is a reason to question a theory. In religion or a belief system, contrary evidence is a reason to question the evidence.

Science is sometimes set up as a "straw man", with conjured images of wrongheaded, lab-coated nerds who would rather look at graphs than listen to music. In a lifetime doing audio research I have yet to find such a person. Rational researchers acknowledge that the subjective experience is "what it is all about". If something sounds "wrong", it *is* wrong. The task is to identify measurements that explain *why* it sounds wrong and, more importantly, to identify those measurements that contribute to it sounding "right", so it can be done again.

In the category of loudspeakers and rooms there is no doubt that differences exist and are clearly audible. Because of this, many reviewers and loudspeaker designers feel that it is not necessary to go to the additional trouble of setting up blind evaluations of loudspeakers. This attitude has been tested.

10.1.1 Why Blind Listening Is Necessary

When I joined Harman International, listening tests were casual affairs. In fact, a widespread practice in the industry has been that a senior executive would have to sign off on a new product, whether they were good listeners or not. Simply sounding "different" on a showroom floor was considered to be an advantage, and sounding louder was a real advantage given that few stores did equal-loudness comparisons. Some managers even had imagined insights into what the purchasing public would be listening for in the upcoming shopping season. In certain markets a prominent product reviewer might be invited to participate (sometimes paid) in a "voicing" session to ensure a good launch of a new product. The final product may or may not have incorporated the advice, and it is unlikely that the deception would ever be discovered.

It was a grossly unscientific process, unjust to the efforts of competent engineers, disrespectful of the existing science, and it contributed to unnecessary variations in product "sounds".

Loudspeaker "voicing". Music composition and arrangement involves voicing to combine musical instruments, notes and chords to achieve specific timbres. Musical instruments themselves are voiced to produce timbres that distinguish the makers. Pianos and organs are voiced in the process of tuning, to achieve a tonal quality that appeals to the tuner or that is more appropriate to the musical repertoire. Recording engineers and musicians are artists creating original sounds, and they are free to introduce colorations and distortions as part of the process. This is all very well, but what has it to do with loudspeakers that are expected to accurately reproduce those tones and timbres?

It shouldn't be necessary if the circle of confusion did not exist, and all professional monitor and consumer loudspeakers were "neutral" in their timbres. However, that is not the case. Recording engineers listen to their loudspeakers and try to anticipate how well their art will "translate" in the consumer world, and loudspeaker manufacturers try to anticipate how well their products will deliver the recorded arts. These people don't talk to each other enough. Consequently, the final stage in loudspeaker development sometimes involves a "voicing" session in which the tonal balance is manipulated to achieve what is hoped to be a satisfactory compromise for a selection of recordings expected to be played by the target customer. There are the "everybody loves bass" voices, the time-tested boom and tizz "happy-face" voices, the polite "slightly depressed upper-midrange voices" (compensating for overly bright close-miked recordings, and strident string tone in some classical recordings), the daringly honest "tell it as it is" neutral voices, and so on. It is a guessing game. Over the years signature sounds were sometimes associated with certain brands. Fortunately, now there is a growing list of manufacturers in both the professional and consumer domains that respect neutrality. It is a welcomed trend.

Until the circle of confusion is eliminated, the guessing game will continue, to the everlasting gratitude of product reviewers, and to the frustration of critical listeners. It is important for consumers to realize that the use of tone controls and equalization are intelligent and practical ways to compensate for inevitable spectral balance variations in recordings. In other words, "voice" the reproduction if and when necessary. Seek out loudspeakers capable of fundamentally neutral, resonance free, reproduction, and acknowledge the fact that no loudspeaker can sound perfectly balanced for all recordings.

Continuing the Harman story, a blind listening facility was set up by the newly established corporate research group, and the engineers were encouraged to incorporate competitive evaluations into the product development routine. Some went along with the plan of striving to be "best in class", using victory in the listening test as the indicator that the design was complete. However, a few engineers regarded the process as an unnecessary interruption of their work, and a challenge to their own judgment. At a certain point it seemed appropriate to conduct a test, a demonstration that there was a problem. It would be based on two listening evaluations that were identical, except one was blind and one was sighted (Toole and Olive, 1994).

Forty listeners participated in a test of their abilities to maintain objectivity in the face of visible information about products. All were Harman employees, so brand loyalty would be a bias in the sighted tests. They were about equally divided between experienced listeners, those who had previously participated in controlled listening tests, and inexperienced, those who had not.

Figure 10.1 shows that in the blind tests, there were two pairs of statistically indistinguishable loudspeakers: the two European "voicings" of the same hardware were rated significantly higher than the other two products, which were also in a statistical tie. In the sighted version of the test, loyal employees gave the big attractive Harman products even higher scores, still tied. However, the little inexpensive sub/sat system dropped significantly in the ratings; apparently its unappealing black plastic exterior overcame employee loyalty. Obviously, something small and inexpensive cannot compete with

something large and stylishly crafted of highly polished wood. The large, attractive competitor improved its rating but not enough to win out over the local product. It all seemed very predictable from the biasing influences. From the Harman perspective, the good news was that two different products were absolutely not necessary for the European marketing regions. (So much for intense arguments that such a sound could not possibly be sold in [pick a country].) In general, seeing the loudspeakers changed what listeners thought they heard.

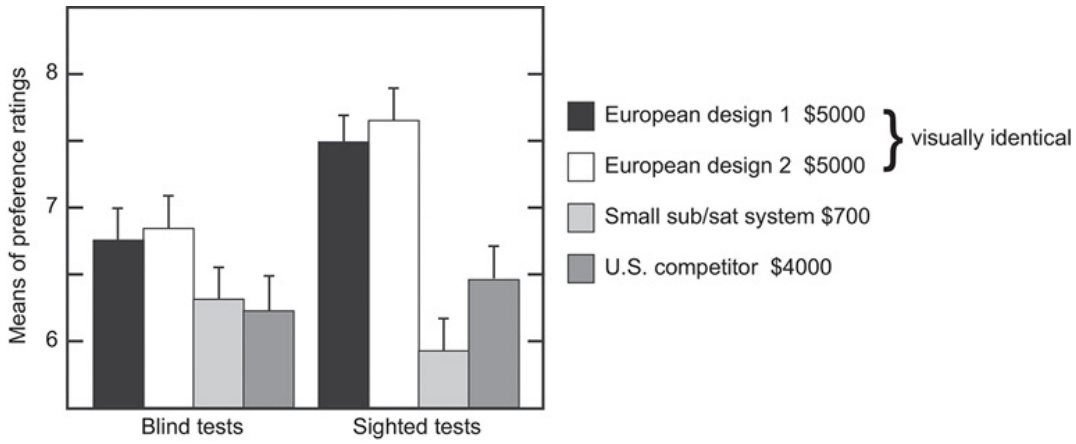


FIGURE 10.1 *The results of blind and sighted listening tests on four loudspeakers, three Harman products and a competing product highly rated in reviews. The two "European design" products differed only in the crossover networks. The sub/sat system was a new design that benefited from excellent engineering, even though it was small and plastic. The lines on top of the vertical bars are the 95% confidence error bars. The large bars must differ by more than this for the difference to be interpreted as significant, with 5% chance of error that they were influenced by uncontrolled variables. From Toole and Olive (1994).* ⏎

The effects of room position on sound quantity and quality at low frequencies are large and well documented. It would be remarkable if these did not reveal themselves in subjective evaluations. This was tested in a second experiment where the loudspeakers were auditioned in two locations that would yield quite different sound signatures.

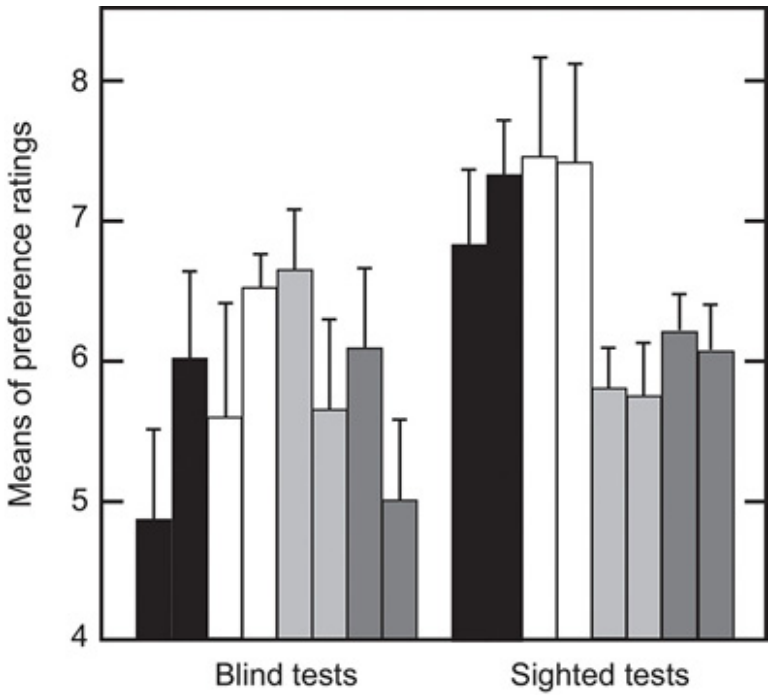


FIGURE 10.2 *A comparison of ratings with loudspeakers in two different locations within the listening room. Adjacent bars of the same shade show results in the two locations for tests conducted under blind and sighted conditions.*

Figure 10.2 shows that in blind evaluations listeners responded to the positional differences: adjacent bars of the same color have significantly different heights, showing the different ratings when the loudspeaker was in position 1 or 2. In contrast, in the sighted tests, there were no significant differences in ratings for the two locations, and the ratings themselves showed the same pattern that was evident in the first experiment (Figure 10.1); listeners obviously recognized the loudspeakers and recalled the ratings they had been given in the first experiment. In addition, some of the error bars in the sighted tests are quite short; no thought is required when you know what you are listening to.

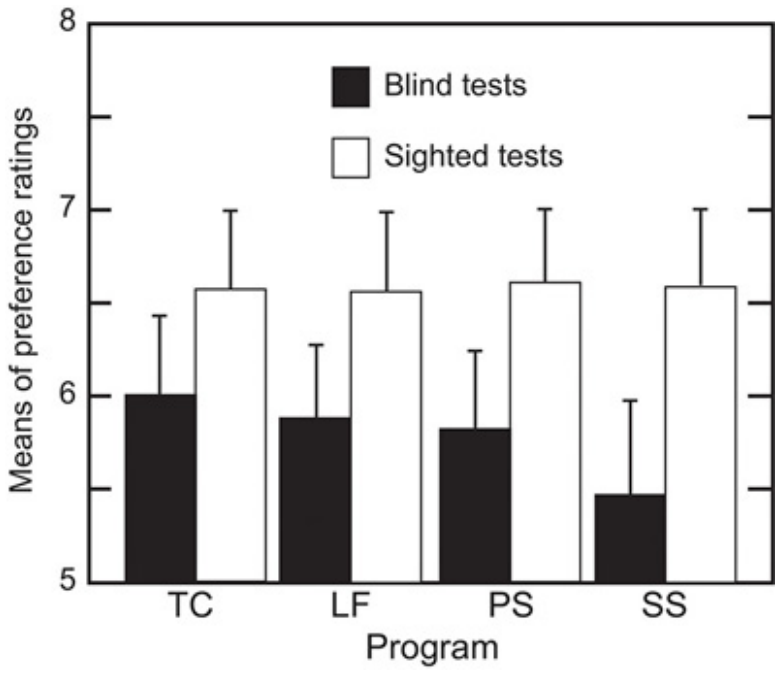


FIGURE 10.3 *Listeners in sighted tests were not responsive to differences caused by different recorded programs.*

It is normal to expect an interaction between the preference ratings and individual programs. In Figure 10.3 this is seen in the results of the blind tests, where the data have been arranged to show declining scores for programs toward the right of the figure. In the sighted versions of the tests, there are no changes, all programs yielded the same ratings. Again, it seems that the listeners had their minds made up by what they saw and were not in a mood to change, even if the sound required it. In a special way they went “deaf”. The effect is not subtle.

Dissecting the data and looking at results for listeners of different genders and levels of experience, Figure 10.4 shows that all groups showed good agreement in their product preferences. Experienced males distinguished themselves by delivering lower scores for all the loudspeakers. This is a common trend among experienced listeners, who are not easily impressed. Otherwise, the pattern of the ratings was very similar to those provided by inexperienced males and females. Over the years, female listeners have consistently done well in listening tests, one reason being that they tend to have closer to normal hearing than males (noisy work environments, or

hobbies?). Lack of experience in both sexes shows up mainly in elevated levels of variability in responses (note the longer error bars), but the responses themselves, when averaged, reveal patterns similar to those of more experienced listeners. With experienced listeners, statistically reliable data can be obtained in less time.

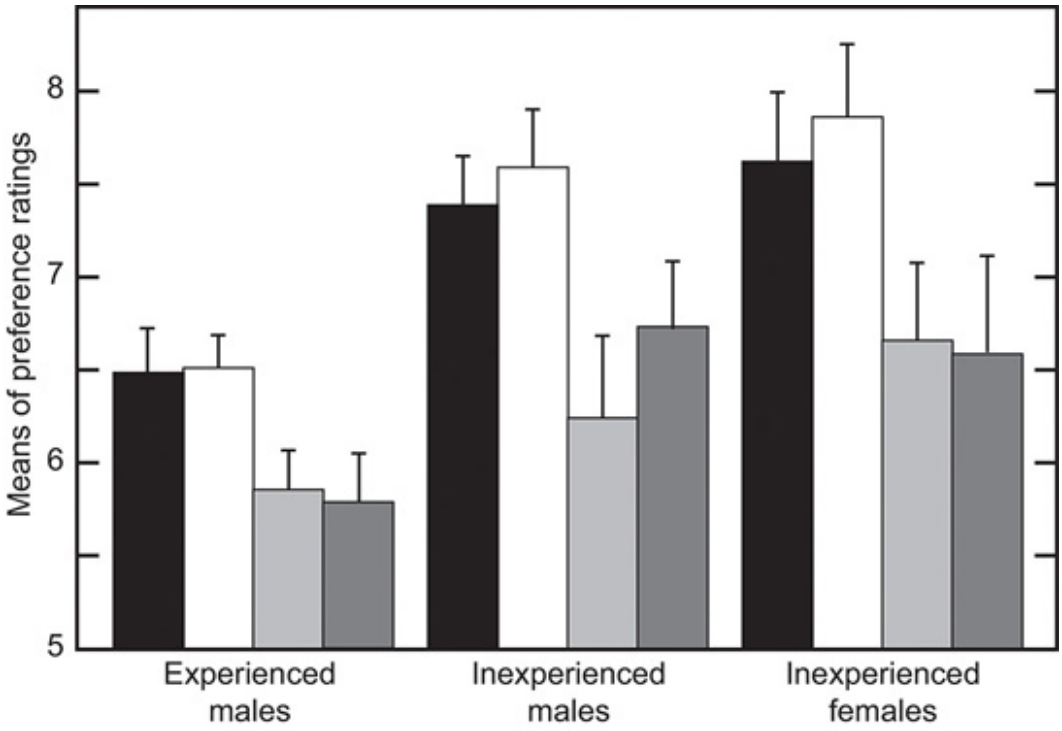


FIGURE 10.4 *The effects of listening experience and sex on subjective ratings of loudspeakers showing results for inexperienced males and females and inexperienced males, there not being any females in the group that had participated in earlier tests.*

Summarizing

- Knowing the identities of the loudspeakers under test changes subjective ratings.
- The ratings change to correspond to believed capabilities of the product, based on price, size, or reputation. This is known as expectation bias.

- The attachment of "perceived" sound quality to the identity of the product is so strong that in sighted tests listeners substantially ignored easily audible problems associated with loudspeaker location in the room and interactions with different programs.
- In sighted tests it is true to say that listeners heard *less* than they did in blind tests.
- Blind testing is essential to obtain meaningful subjective opinions on the sound quality of loudspeakers. There is a placebo effect in audio.
- The good news is that if the appropriate controls are in place, experienced and inexperienced listeners of both genders can deliver useful opinions. Inexperienced listeners take longer, delivering more variable ratings which therefore require more repetitions to produce the same confidence levels in their ratings.

10.1.2 Hearing Ability and Listening Performance

Opinions of sound quality are based on what we hear. If hearing is degraded, we simply don't hear everything, what we do hear is not normal, and opinions will be affected. Hearing loss is common, and it can happen suddenly or slowly develop over the years. It is something that many people don't talk about, especially those for whom it matters, like musicians, recording engineers and reviewers. The symptoms vary greatly from person to person. In audio, opinions from such persons may or may not be relevant to others. As I sit here with my never-ending tinnitus, I don't (yet) need hearing aids, I greatly enjoy music and movies, and I have opinions, but they are mine alone. Several of our friends have significant hearing problems, some from military duty, others from occupational and/or recreational noise exposures, and some from medical issues. Hearing aids can improve verbal communication, but "hi-fi" hearing is gone forever. Adding to the insult, loud sounds can be uncomfortable, distorted or even painful.

These are obvious hearing problems, but long before one is aware of a problem, and even when your audiologist declares that you have "normal" hearing, there can be significant changes in one's ability to form opinions

about sound quality. For audiophiles this may be no more than a nuisance, even if they are aware of it. For audio professionals it is a handicap—one that they often are not aware of, but that obviously has an effect on their product.

The first clear evidence of the issue with hearing came in 1982 during an extensive set of loudspeaker evaluations conducted for the Canadian Broadcasting Corporation (CBC), in which many of the participants were audio professionals: recording engineers, producers, musicians. Unfortunately, for them, hearing loss is an occupational hazard. During analysis of the data, it was clear that some listeners delivered remarkably consistent overall sound-quality ratings (then called fidelity ratings, reported on a scale ranging 0 to 10) over numerous repeated presentations of the same loudspeaker. Others were less good, and still others were extremely variable, liking a product in the morning and disliking it in the afternoon, and each time writing a clear explanation why. The cause was not hard to find. Separating listeners according to their audiometric performances, it was apparent that those listeners with hearing levels closest to the norm (0 dB) had the smallest variations in their judgments. Figure 10.5 shows examples of the results. Surprisingly, it was not the high-frequency hearing level that correlated with the judgment variability but that at frequencies at or below 1 kHz, people who also have high-frequency loss. Bech (1989) confirmed the trend.

Noise-induced hearing loss is characterized by elevated thresholds around 4 kHz. Presbycusis, the hearing deterioration that occurs with age, starts at the highest frequencies and progresses downward. These data showed that by itself, high-frequency hearing loss did not correlate well with trends in judgment variability (see Figure 10.5(b)). Instead, it was hearing level at lower frequencies that showed the correlation (Figure 10.5(a)). Some listeners with high-frequency loss had normal hearing at lower frequencies, but all listeners with low-frequency hearing loss also had loss at high frequencies—in other words, it was a broadband problem. We now know that elevated hearing thresholds are often accompanied by reduced ability to separate sounds in space, which would include separating the listening room from the sound of the loudspeaker, which adds significant difficulty to the task. This is very likely a broadband effect; see Chapter 16. This suggests that eliminating room reflections might be advantageous for those with hearing loss.

Illustrating the situation more clearly is Figure 10.6, which displays the hearing performance of the author as a function of age, and a collection of threshold measurements for the high-variability listeners in Figure 10.5. The hearing threshold data are shown superimposed on the family of equal-loudness contours, the sound levels at which different pure tones are judged to be equally loud. The bottom curve is the hearing threshold, below which nothing is heard.

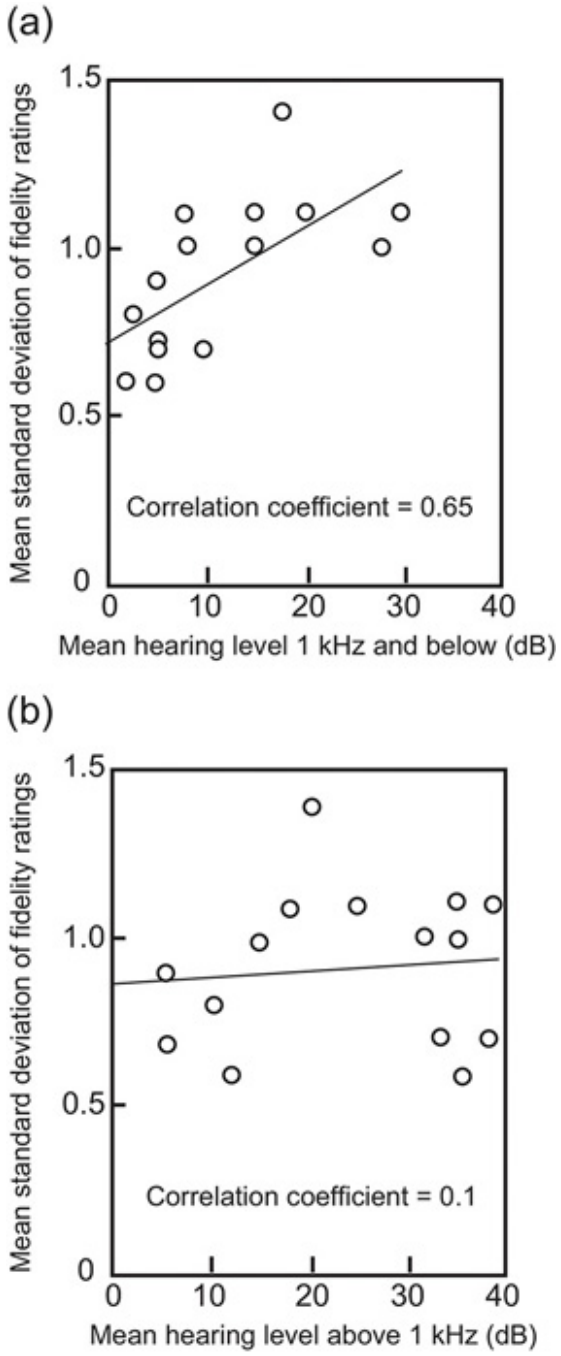


FIGURE 10.5 *Variability in "fidelity" rating judgments (mean standard deviation) as a function of hearing level (0 dB indicates normal hearing, and higher levels indicate that hearing thresholds have been elevated by that amount). Sounds at lower levels are not audible by those persons. Audiometric measurements were made at the standard frequencies of 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz and 8 kHz. Hearing levels are the amounts by which the measured*

thresholds for the listeners are higher than the statistical normal hearing threshold. It is a measure of hearing loss. In (a) hearing levels at 1 kHz and below were averaged. In (b) hearing levels above 1 kHz were averaged. From Toole (1985, figure 8). ⏎

My thresholds at age 30 were slightly below (better than) the statistical average. With age, and acoustical abuses of various kinds, the hearing level rose gradually, most dramatically at high frequencies. Even so, my deterioration was less than the population average presbycusis for age 60. My broadband auditory dynamic range was degraded by only 10 dB or so over those years, except that high frequencies disappeared at a higher rate.

In contrast, the professional audio people whose audiometric data are shown in Figure 10.6(b) were unable to hear large portions of the lower part of the dynamic range. Without the ability to hear small acoustical details in music and voices (good things) or low-level noises and distortions (bad things) these people were handicapped in their abilities to make good or consistent judgments of sound quality, and this was revealed in their inconsistent ratings of products. It did not prevent them from having opinions, deriving pleasure from music or being able to write articulate dissertations on what they heard, but their opinions were not the same as those of normal hearing persons. Consequently, future listeners underwent audiometric screening, and those with broadband hearing levels in excess of about 20 dB were discouraged from participating.

Many of these hearing-impaired listeners were active recording engineers and producers, responsible for creating the programs we listen to. They are not alone in the industry. Martinez and Gilman (1988) report results from an audiometric surveys of AES members, including many recording engineers. It is evident that hearing loss is not a new topic. Music-induced hearing loss (MIHL) is a well-recognized problem. Jamieson (2015) provides an excellent perspective on the present situation, including surveys of industry participants, how they perceive and deal with their own problems, and discussions of how some of them employ hearing protection at least part of the time. Recording engineers are in a difficult situation, needing to listen for extended periods to do their jobs, knowing that there is a high risk of damage to their hearing. The

survey indicates that hearing loss has forced some to change occupations. Some know they have hearing loss but don't think it affects their work, believing that they can compensate.

This reality suggests that hearing performance should have been included in the "circle of confusion" (Figure 1.8), because it is a factor.

All this emphasis on normal hearing for subjects in listening tests seems to imply that a criterion excluding listeners with greater than 15–20 dB hearing level may be elitist. According to USPHS data, about 75% of the adult population should qualify. However, there is some concern that the upcoming generations may not fare so well because of widespread exposure to high sound levels through headphones and other noisy recreational activities.

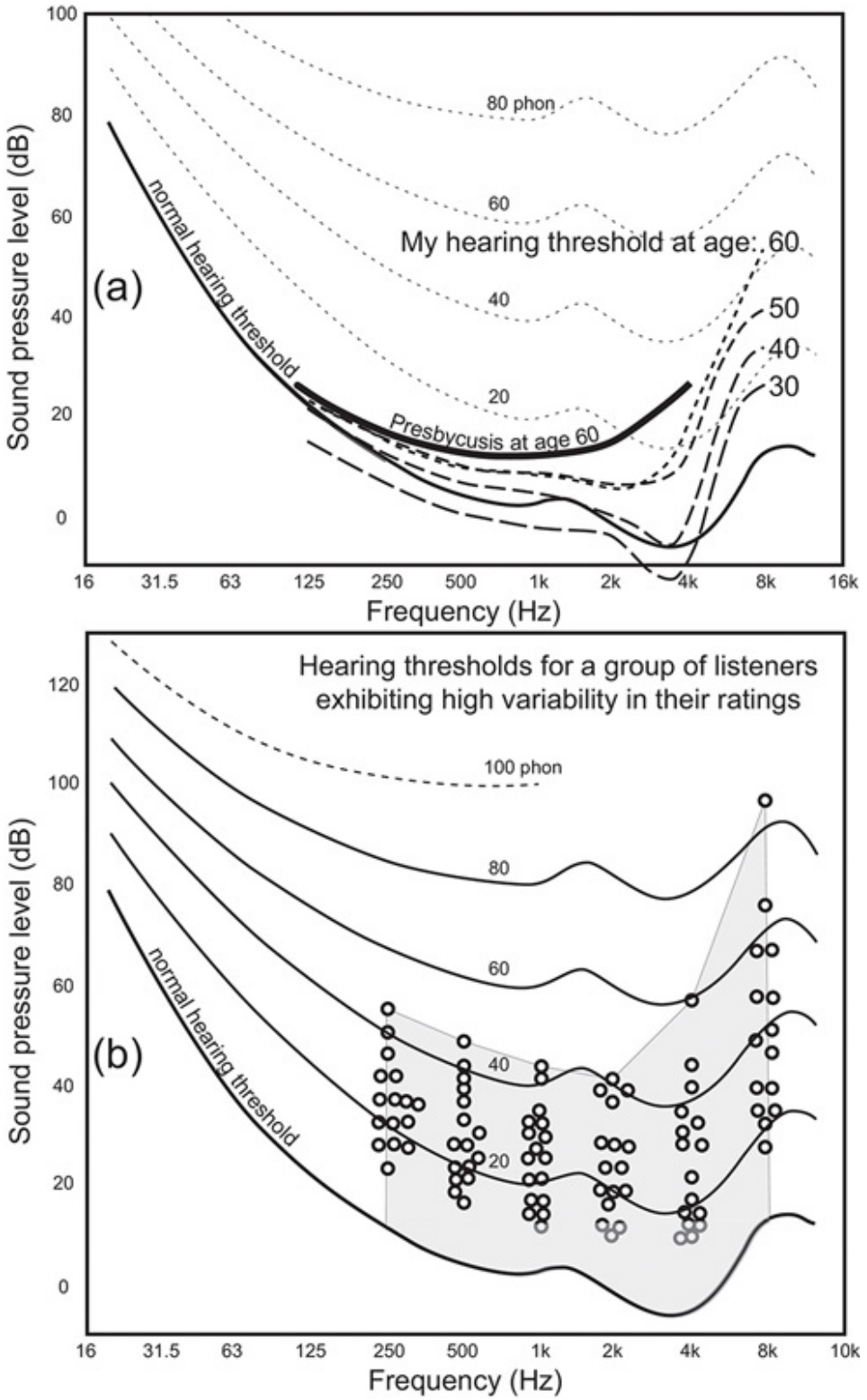


FIGURE 10.6 *(a) A history of the author's hearing threshold compared with the statistical normal hearing threshold and with the normally anticipated hearing loss as a function of age (presbycusis).*

(b) The hearing thresholds for listeners who had high variations in repeated sound-quality ratings of loudspeakers. Adapted from Toole, 1985, figure 7. Equal loudness contours are from ISO 226 (2003). *© ISO. This material is reproduced from ISO 226 with permission of the American National Standards Institute (ANSI) on behalf of the International Organization for Standardization (ISO). No part of this material may be copied or reproduced in any form, electronic retrieval system or otherwise or made available on the Internet, a public network, by satellite or otherwise without the prior written consent of the ANSI. Copies of this standard may be purchased from ANSI, 25 West 43rd St., New York, NY 10036, http://webstore.ansi.org.* ⏎

Hearing loss occurs as a result of age and accumulated noise exposure over the years. Figure 10.7(a) shows that in terms of our ability to make reliable judgments of sound quality, we do not age gracefully. Two of the data points indicate that young persons are not immune to hearing loss. It is a problem traceable to both age and regular exposure to loud sounds as is confirmed by Figure 10.7(b), an audiometric survey of Audio Engineering Society convention attendees, all of whom have a high probability of exposure to high sound levels. In my younger years, I was an excellent listener. Listening tests as they are done now track not only the performance of loudspeakers but of listeners—the metric shown in Figure 10.9. About age 60 it was clear that it was time to retire from the active listening panel. Variability had climbed and I found it to be a noticeably more difficult task. It is a younger person's pursuit. My extensive experience and knowledge was no advantage. Music is still a great pleasure, but my opinions are now my own. When graybeards expound on the relative merits of audio products, they may or may not be relevant. But be polite—the egos are still intact.

Figure 10.8 is a revealing display of how listeners with different hearing abilities judge sound quality. Listeners evaluated four loudspeakers in a monophonic multiple-comparison blind test. Each dot is the mean of several judgments on each of the products, made by one listener. To illustrate the effect of hearing level on ratings, listeners were grouped according to the

variability in their judgments: those who exhibited standard deviations below 0.5 scale unit and those above that. The low-variability listener results are shown in the white areas and the high-variability results are in the shaded areas.

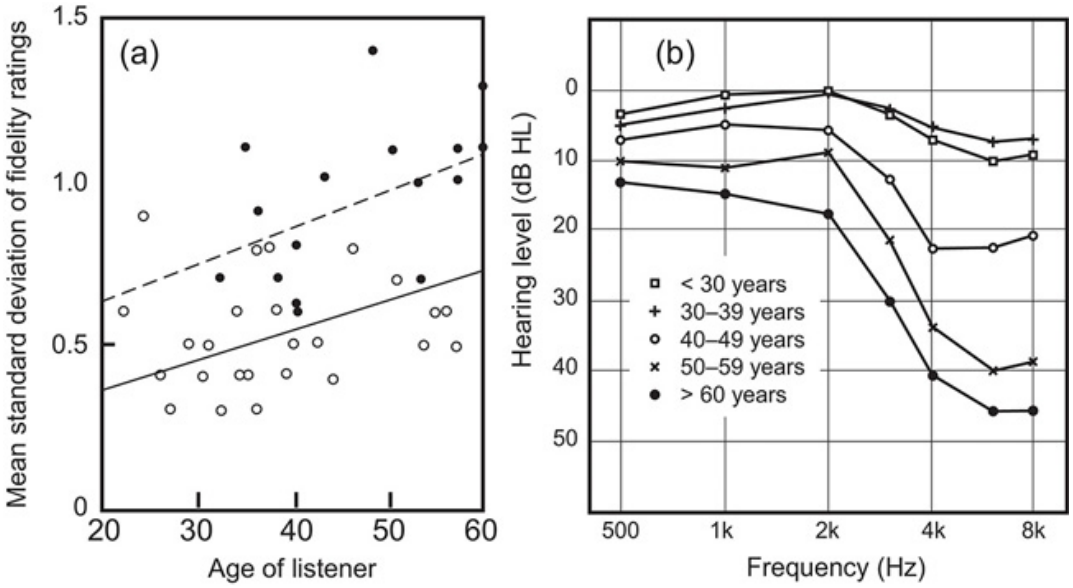


FIGURE 10.7 *(a) Variations in "fidelity" ratings vs. the age of the listener. The results are for two sets of experiments. The black dots and dashed regression line represent an early test in which all controls were not in place. The open circles and solid regression line are from well-controlled tests and therefore show lower variability. However, both sets of data display the same overall trend of increasing variability with increasing age. From Toole, 1985, figure 9. Hearing loss from age and noise exposure are common among audio professionals. (b) shows audiometric measurements on 229 attendees at a 1986 AES convention. All persons were involved with the audio industry, from recording engineers through support staff and vendors. 0 dB represents normal hearing, the hearing threshold curve in Figure 10.6. Here the hearing level is presented in the conventional audiometric style, as a "loss", instead of showing actual elevated threshold levels as in Figure 10.6. These results show the combined effects of age and exposure to high sound levels,*

and reveal high-frequency loss as well as low-frequency loss. From Martinez and Gilman (1988). ⏎

In Figure 10.8, it is seen that fidelity ratings for loudspeakers D and U are closely grouped by both sets of listeners, but the ratings in the shaded areas are simply lower. Things change for loudspeakers V and X, where the close groupings of the low-variability listeners change to widely dispersed ratings by the high-variability listeners. This is a case where an averaged rating does not reveal what is happening. Listener ratings simply dispersed to cover the available range of values; some listeners thought it was not very good (fidelity ratings below 6), whereas others thought it was among the best loudspeakers they had ever heard (fidelity ratings above 8). Listeners simply exhibited strongly individualistic opinions. Hearing loss is very likely involved. Be careful whose opinions you trust.

Other investigations agree. Bech (1992) observed that hearing levels of listeners should not exceed 15 dB at any audiometric frequency and that training is essential. He noted that most subjects reached a plateau of performance after only four training sessions. At that point, the test statistic F_L should be used to identify the best listeners.

Olive (2003) compiled data on 268 listeners and found no important differences between the ratings of carefully selected (for normal hearing) and trained (to be able to recognize resonances) listeners and those from several other backgrounds, some in audio, some not, some with listening experience, some with none. There were, as shown in Figure 10.9, huge differences in the variability and scaling of the ratings revealing how consistently listeners rated the products in repeated exposures, and how confidently they were able to separate the ratings. Selection and training of listeners have substantial benefits in time savings. Rumsey et al. (2005) also found strong similarities in ratings of audio quality, anticipating only a 10% error in predicting ratings of naïve listeners from those of experienced listeners.

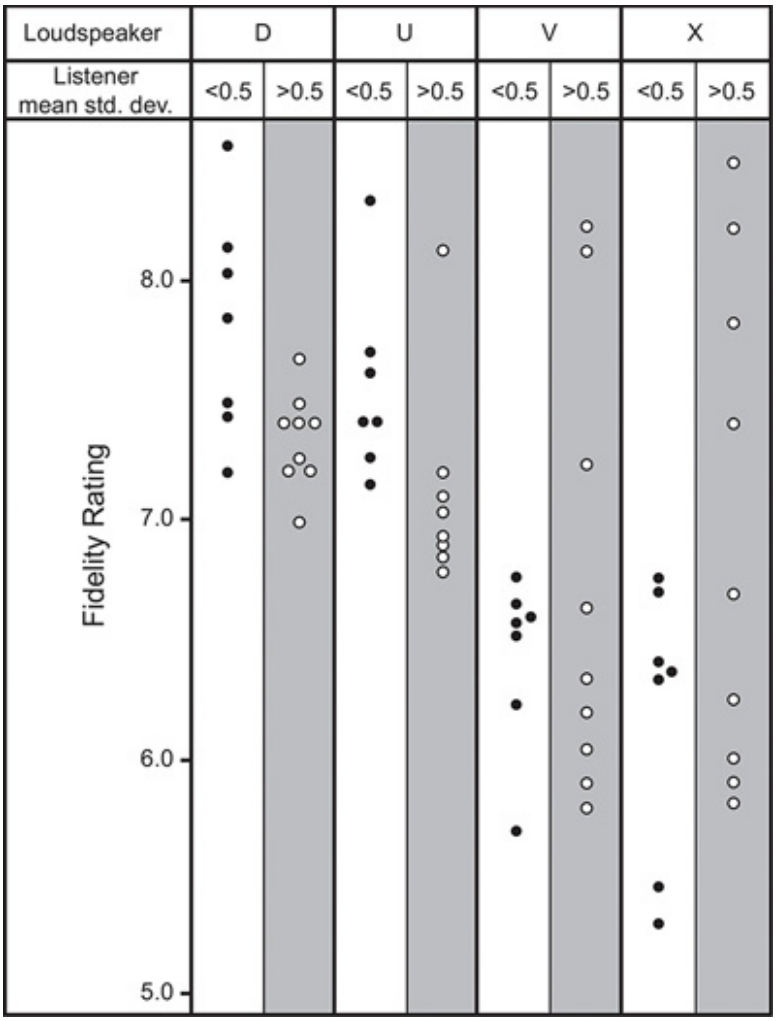


FIGURE 10.8 *Averaged sound quality rating judgments for listeners classified according to the variability in their ratings. The small variability/more normal hearing listener results are shown in the white columns. The large variability/less-than-normal hearing listener results are shown in the shaded columns. For interest: D was a PSB Passif II, U was a Luxman LX-105, V was an Acoustic Research AR58s, and X was a Quad ESL-63. From Toole, 1985, figure 16.*

The good news for the audio industry is that if a loudspeaker is well designed, ordinary customers may recognize it—if given a reasonable opportunity to judge. The pity is that there are few opportunities for consumers to make valid listening comparisons, and the author is not aware of

an independent source of such *unbiased* listening test data for customers to go to for help in making purchasing decisions. All of this underlines the value of having trustworthy and meaningful technical data to assist in making decisions about loudspeakers.

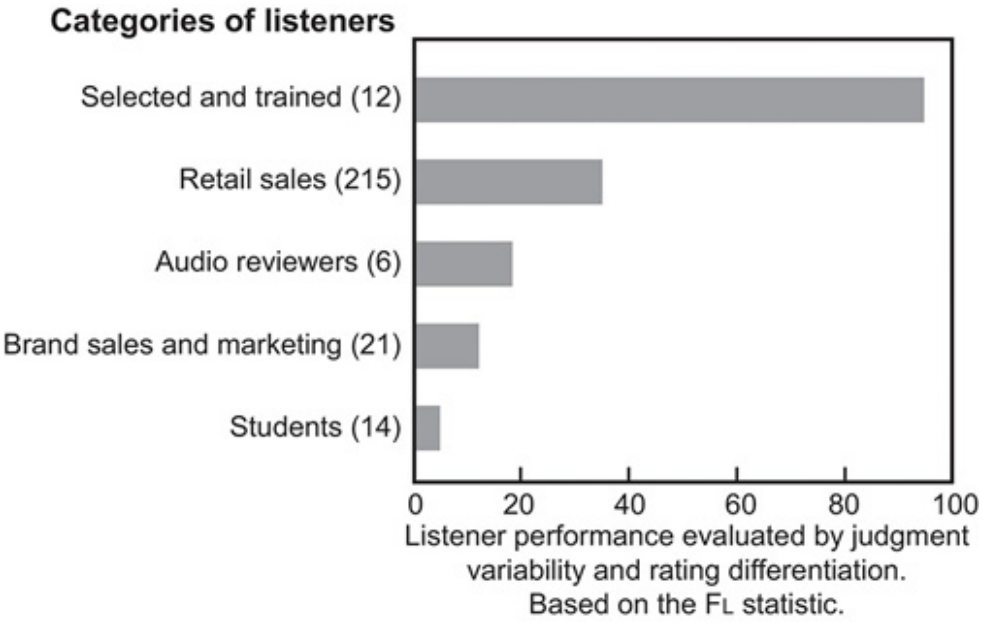


FIGURE 10.9 *A measure of listener performance that is related to the consistency of repeated judgments as well as the strength of rating differentiation of products having different sound quality levels (based on the FL statistic). Twelve selected and trained listeners rated around 95 on a scale of 100. Two hundred fifteen retail sales people rated around 40, 6 audio reviewers rated about 20, 21 brand sales and marketing people rated about 12, and 14 students rated about 4. In spite of these differences, averaged product ratings were very similar for all groups. From Olive, 2003.*

10.2 LOUDSPEAKER REVIEWING IS A CHALLENGE FEW CAN MEET

Product reviews are extremely popular. Many people are looking for guidance. Print and online audio magazines have provided it over the years, and now the

internet has taken over. Reviewing is a regular activity for some people whose names are familiar from reading or viewing the various sources in which their works appear. Enthusiasts get to know them, to develop a trust in their opinions, or not. A few of them take the job seriously, studying the science and technology behind the products, and, when listening, taking care to minimize bias, while recognizing that it is not completely possible. The majority of these reviews employ the sighted "take it home and listen to it" method, which we know to be susceptible to bias and adaptation (see Section 1.4).

Few publications have the budget to make accurate acoustical measurements and do properly controlled listening tests. In the United States, Consumers Union, and their publication *Consumer Reports*, is such an organization, and for many years they published reviews of loudspeakers based on anechoic measurements elaborately processed in a scientific-sounding manner. Their ratings were greatly influential for the general public, but there were many scientific and technical people, the author included, who questioned their methods and told them so. As will be discussed in the following chapter, research done at Harman by Sean Olive proved them wrong, and they ceased reviewing loudspeakers. This was a responsible act on their part, but it took far too many years for them to acknowledge the errors in their methods.

As discussed in Section 2.9, measurements have changed the loudspeaker industry and, while expensive, at least two reviewers have equipped themselves with Klippel near-field scanners to make trustworthy comprehensive anechoic measurements (e.g. www.audiosciencereview.com, www.erinsaudiocorner.com). Other reviewers purchase anechoic data (e.g. www.soundstagenetwork.com). Others make a credible effort to acquire trustworthy data outdoors (e.g. www.audioholics.com, www.stereophile.com). I see evidence of others on YouTube, but I have not followed them long enough to make a judgment, and there are likely to be more that I am not aware of, in which case congratulations and my apologies.

When meaningful measurements are incorporated in a review, there can be insightful commentary that relates to demonstrable product characteristics. The subjective comments included in these documents are not definitive, but

the serious reviewers know it, and do what they can to minimize bias, like listening before doing measurements. The measured data is hard evidence that sometimes challenges the subjective impressions, especially as the listening rooms are different, programs are different, and so on. These reviews take time, cost money and are documents of lasting value, the measurements are content I have returned to several times over the years and these reviewers have my gratitude.

In the absence of useful anechoic or simulated-anechoic technical evidence —room curves do not qualify—purely subjective reviews are demonstrably inconsistent. Many of them are really telling us what they heard when listening to specific recordings. Are they record reviews or loudspeaker reviews? When they start writing essays on the "sound" of speaker and interconnect wires, power cords, stands, vibration isolators, spikes, and most electronics, without measurements or double-blind evaluations, one knows that they are in the entertainment business, supporting their advertisers or themselves.

But "reviews" are everywhere now, liberated by the internet. Anyone with an opinion and a keyboard is free to offer anything from brief, pithy comments to long "record review" type essays on what they "heard" through various audio components. It is easy to understand the motivations. For ardent hobbyists it is a way to share their enthusiasm for audio and for some products they like, maybe even to earn some money. But, in the end, their purely subjective opinions are of little value because the listening performed by most of them violates the most basic rules of good practice for eliminating bias. There may be articulate descriptions of the physical product, discussing real or imagined virtues in the transducers, cabinet construction, connectors, etc., but no proofs of performance. All probably mean well, some know their music and appear to be decent record reviewers, and a few are entertaining. But their opinions about the sound of the loudspeakers themselves are personal, not for sharing. As with most works of fiction, smile, close the book, and move on to the next one.

There are opportunities for independent parties to do valid, equal-loudness, double-blind subjective evaluations with trained listeners having normal hearing. It requires a room, positional substitution apparatus, time, money, and

discipline. Many loudspeaker manufacturers don't do it properly, but most have measurement capability. As the research reported in this book demonstrates, having both is very powerful. SenseLab, a Danish company, performs subjective evaluations on products, and sells rotating platforms to enable position-substitution loudspeaker tests (https://forcetechnology.com/en/all-industry-facilities/acoustics-noise-sound-quality-laboratories/senselab-sound-quality-laboratories).

10.3 DESIGNING LISTENING TESTS

Subjective evaluations are done for a variety of purposes, from evaluating products to decide whether they are worthy of being manufactured, to selecting loudspeakers to purchase and take home, to judging their performance for a produce review to be published. All need to be at least somewhat controlled to yield useful, reliable results, but not everyone works with the same budget or time scale. Knowing what the key variables are can help in the optimum design of tests for different purposes. Much has been written on the topic (Toole, 1982, 1990; Toole and Olive, 2001; Bech and Zacharov, 2006). The following is a summary.

10.3.1 Is Stress a Factor? Relax

Purely subjective reviewers have criticized double blind testing, arguing that it puts extraordinary stress on listeners, thereby preventing their full concentration on the task. This is presumed to explain why people in blind tests don't hear the same things that show up in their casual, live-with-the-speaker, listening sessions. It could be true, but years of doing these tests with numerous listeners have shown that, after a brief introductory period, they simply settle into the task. Regular listeners have no anxiety. Once they see that their opinions are not aberrant, that they are meaningfully related to measured loudspeaker performance, they become confident "measuring instruments". Some enjoy the challenge.

Years ago, at the NRCC, we contracted to perform anechoic measurements and to conduct blind loudspeaker comparisons for Canadian audio magazines. The journalists traveled to Ottawa, arriving the night before the tests to give

their ears a chance to rest, and spent a day or more doing double blind subjective evaluations before they got to see the measurements on the products they submitted. On one occasion, due to unforeseen circumstances, the special acoustically transparent, visually opaque, screen was not available —it was on a truck 2 hours away. Rather than waste time, it was decided to start the tests without the screen—sighted. A few rounds into the randomized, repetitive tests our customers stopped the proceedings—they felt that seeing the loudspeakers prevented them from being objective—they were stressed! They waited for the truck to arrive.

Those of us who have done many of these tests agree. In the end it is simpler if there is only one task—to judge the sound. There is no need to wonder if one is being influenced by what is seen, noting that certain familiar brand names might not be performing as anticipated, thinking you might be wrong, and so on. In the multiple comparison tests conducted for these customers, we always added one or more products to the test population that are not known to the listeners to discourage guessing. As a result, listeners gave up trying to guess what they were listening to—it is another distraction that often proves to be wrong.

Additional criticisms from subjectivists had to do with notions that the sound samples might be too short for critical analysis, and that a time limit can be stressful. The automated positional substitution apparatus at Harman permitted tests for single listeners, who could participate in a test when it was convenient. In these tests, the listener selects which of the three or four unidentified loudspeakers is to be heard at any time, and there is no time limit —switch at will, and take as long as is needed. The musical excerpts are short but repeated in nicely segued loops to allow for detailed examination. Stressful? It is true that the musical pleasure is lost, but the power of the test is increased, and the duration reduced. We are not enjoying the music or doing record reviews; we are evaluating loudspeakers.

10.3.2 Stereo vs. Monophonic Listening

Stereo is the default format for most program material. So, should not all loudspeaker evaluations be done in stereo? In 1985 I decided to examine the effects of loudspeaker directivity and adjacent side-wall reflections, and the

evaluations of sound quality and spatial quality were done in both stereo and mono (Toole, 1985, described in Section 4.4). Because stereo imaging and soundstage issues were involved, listeners were interrogated on many aspects of spatial and directional interest. To our great surprise, listeners had strong opinions about spatial qualities when listening in mono. Not only that, but they were much more strongly opinionated about the relative sound quality merits of the loudspeakers in the mono tests than in the stereo tests. Early loudspeaker comparisons were done in mono for simplicity in making comparisons, but with this new information monophonic evaluations continued.

Naturally there were doubters, and over the years mono vs. stereo tests have been performed. Each time the loudspeaker that won the mono tests also won the stereo tests, but lower rated loudspeakers were given higher ratings. In stereo everything tended to sound better. See Section 4.4 for details. Chapter 5 looks deeper into how we make judgments of sound quality in different circumstances. It concluded that in stereo and multichannel sound recordings include enough reflected sound information about real or synthesized recording environments to confuse the eye/ear perceptual system when all channels are active.

A persuasive justification for monophonic evaluations is the reality that in movies and television programs the center channel does most of the work—a monophonic sound source. The center channel is arguably the most important loudspeaker in a multichannel system. In popular music and jazz stereo recordings, images are often hard-panned to the left or right loudspeakers, again monophonic sources, and the phantom center image is the product of double mono, identical sounds emerging from L and R loudspeakers.

Finally, the data of Figure 4.12 adds more strength to the argument that monophonic listening has merits. It shows that only those listeners with hearing levels very close to the statistical normal level (0 dB) are able to perform similarly well in both stereo and mono tests. Even a modest elevation in hearing level causes judgment variations in stereo tests to rise dramatically, yielding less-trustworthy data. Age and hearing loss affects one's ability to separate multiple sounds in space. We may notice this in a crowd, or in a restaurant setting but we may not be aware of the disability when listening to a

stereo soundstage. The directional and spatial complexity of stereo appears to interfere with our ability to isolate the sound of the loudspeakers, see Sections 5.7 and 5.8.

For program material, some people insist on using one of a stereo pair of channels, but summing the channels is possible with many stereo programs; listen and find those that are mono compatible. Since complexity of the program is a positive attribute of test music, abandoning a channel is counterproductive. In these tests one is not judging artistic merit, the music is a test signal.

Sections 4.4 and 4.5 were summarized with this:

> Finally, advice to loudspeaker users: evaluate loudspeakers in single loudspeaker monophonic comparisons (to find out what you have). Then, if they are good, demonstrate them in stereo or multichannel (to impress everybody). Choose the recordings carefully—they are a significant factor in how pleased you will be.

10.3.3 The Listening Room

As discussed in earlier chapters, adaptation to the listening room is a factor in what we are able to hear in perceptions of loudspeaker sound quality. Below the transition/Schroeder frequency, resonances in small rooms have a massive influence on what we hear. They add timbral colorations and are the acoustic "signature" of a small room, not the large spaces we want to hear in recordings. There are also effects at higher frequencies, in the reflection of off-axis sounds related to loudspeaker directivity. Therefore, if comparing loudspeakers, listen to all of them in the same room, taking time to let the listeners adapt to the acoustical setting. Listeners have the ability to adapt to many aspects of room acoustics in multiple-loudspeaker listening comparisons, as long as there are no extreme problems. As shown in Sections 5.4 and 5.5, when comparisons are done in the same room, large differences in rooms make little difference to the relative sound quality ratings. Listeners

appear to be able to normalize the effects of different rooms, at least up to a point.

That does not mean that there are *no* effects. Room resonances are there, modifying the timbre of the loudspeakers, causing listeners in different locations to hear different bass sounds, and causing loudspeakers in different locations to deliver different bass sounds to each listener, wherever they are seated. As has been mentioned, low-frequency performance accounts for roughly 30% of overall sound quality ratings, so it is necessary to pay attention to it.

10.3.4 Solutions for the Loudspeaker Position Problem

Section 1.4 discussed listening test methods, concluding that the multiple-loudspeaker technique, three or four loudspeakers at a time, is the most efficient and possibly the most effective way to identify the timbral differences among products. The following assumes that this method will be used.

Ideally, use positional substitution: bring the active loudspeakers to the same position in the room so that the coupling to the room modes is constant, as are adjacent-boundary effects and wall reflections. This can be expensive, and may require a dedicated facility, something beyond the means of hobbyists, or even audio publications and their reviewers. But, with the expenditure of some time and effort, simpler methods can yield very reliable results.

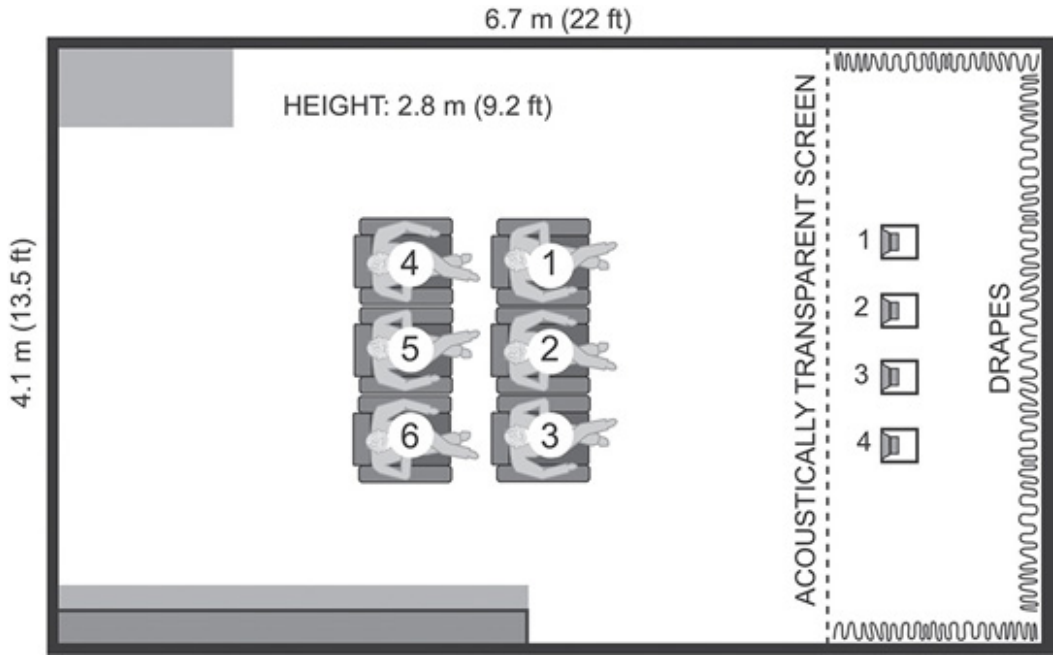


FIGURE 10.10 *The original listening room at the NRCC, ca. 1982, set up for monophonic comparisons of four loudspeakers. In the course of the evaluations the loudspeakers and listeners would be randomly rearranged to average out the strong influences of standing waves in the room. See Figure 2.1 for a photograph of a common listening setup.* ⏎

Figure 10.10 shows the listening room used in my original research at the NRCC, in the 1980s. In this, three or four loudspeakers were located at the numbered locations, and some number of listeners occupied the seats. Rigorous tests used one listener at a time. After each sequence of music, during which listener(s) made notes and decided on "fidelity" ratings, the loudspeakers were randomly repositioned, and the listeners took different seats. This process was repeated until all loudspeakers had been heard in all locations. It was time-consuming, but it reduced the biasing influence of room position, which in some cases could be the deciding factor. See Figure 10.2 for evidence of the importance of position.

The setup needs to be examined using steady-state measurements from each loudspeaker location to each listener location, and positional adjustments made so as to avoid accidental contamination by one or more aggressive room resonances. The appendix of Toole (1982) describes the process for one room,

and Olive et al. (1998), Olive (2009) and McMullin et al. (2015) document others. The room cannot be eliminated, but it should be a relatively constant and innocuous, factor.

If the result is to be an evaluation relevant to ordinary audiophiles or consumers, deliberate acoustical treatment must be moderated. Imitating an "average" domestic listening room is a good objective and the most contentious decision is likely to be what to do with the first side wall reflection. My initial decision, many years ago, was to let it reflect—a flat untreated surface, as happens in many domestic rooms and retail demonstrations—so that the off-axis behavior of the loudspeaker would be revealed. This is a demanding test, as it probably should be for loudspeaker evaluations intended to apply to a general population. But, for one's personal recreational listening there are other factors to consider, as is discussed in Section 8.5.5, there are options.

The results of these simple experiments were good enough to reveal the essential facts published in the benchmark Toole (1985, 1986) papers, and to provide a solid foundation for the research that followed. Using a similar setup for listening tests is labor and time intensive, but the results are meaningful.

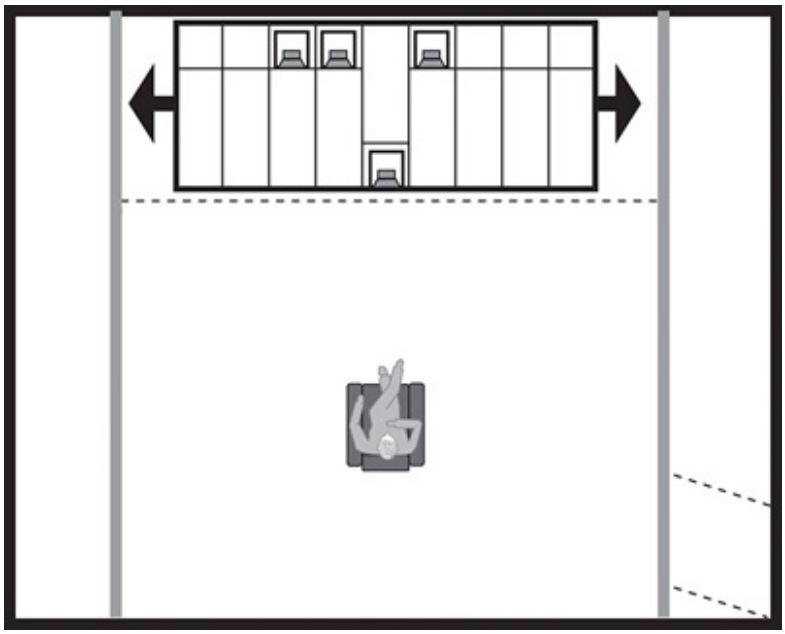


FIGURE 10.11 *The Harman International Multichannel Listening Laboratory (MLL) showing the side walls in their revised, closer to*

center, locations. The programmable mover can compare up to four single loudspeakers at one time, up to four stereo pairs, or up to three L, C, R combinations. The single loudspeakers can be compared at L, C or R locations at varied distances from the side walls. The distance from the front wall is also a controlled variable. ⏎

When the research group was established at Harman International, one of its purposes was to provide double-blind listening tests to evaluate loudspeakers under development. In an engineering environment it was important that these data be available quickly, so as not to delay the development of products. We were fortunate to have a budget that permitted building a custom listening facility, a soundproofed room, including a positional substitution mechanism that would allow for rapid comparisons of mono, stereo or L, C, R multichannel loudspeaker configurations. It was called the Multichannel Listening Laboratory (MLL) (Olive et al., 1998).

The positional substitution process was automated using computer controlled pneumatic rams to move the entire platform left and right, and to move the desired loudspeaker(s) to a predetermined forward location, storing the silent loudspeakers at the rear. See Figures 10.11 and 10.12. The amount of forward movement is programmable, and stereo and L, C, R arrangements can be compared. The positional exchanges take about 3 s, and take place quietly. The single listener controls the process, choosing the active loudspeaker(s) (there was a displayed identity code) and deciding when it is time to move on to the next randomly chosen musical selection. When the music changes the coded identities are also randomly changed, so in effect it becomes a new experiment. There is no time limit, the listener has to be satisfied before moving on. All functions are computer randomized to make it a true double-blind test. There is a video camera to ensure that no cheating (looking behind the screen) takes place.

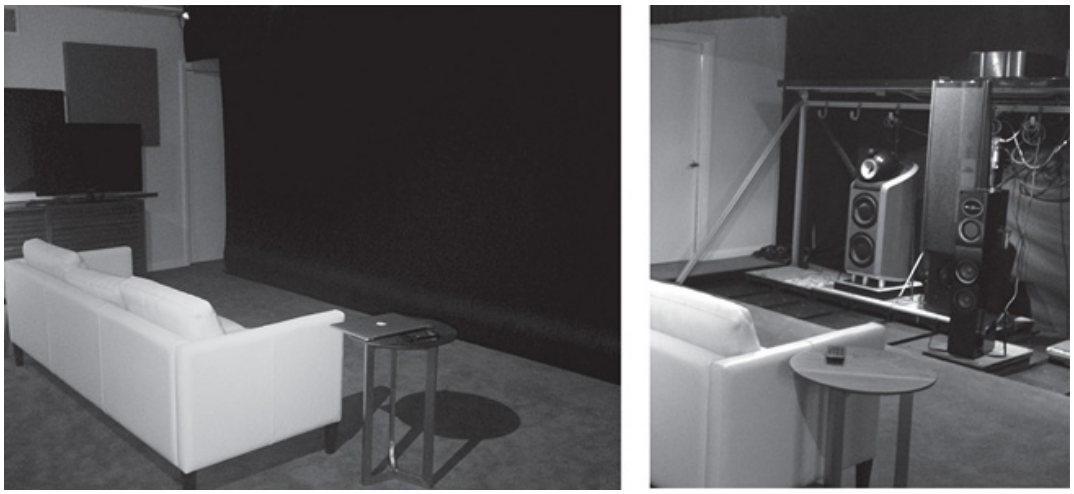


FIGURE 10.12 *Photographs of the MLL with and without the acoustically transparent curtain, showing three loudspeakers under evaluation, one of them in the forward, active, location.*

This process was new to the Harman engineers. They had questions, and there were doubters. However, when the results started being revealed, and the engineers participated in the double-blind evaluations of their own products, they became enthusiastic users. They could see the correlations between their measurements and listener ratings. Winning the double-blind comparison with prominent competing products became the challenge. All this subjective data was accumulated along with associated anechoic measurements, adding to the scientific database leading to the subjective/objective correlations in the next chapter.

There are other ingenious ways to achieve positional substitution for in- or on-wall loudspeakers, sound bars, etc. If these tests are to be a way of life, any are worth considering because of the excellent results they yield and the time savings involved. If none of this is possible, a variation of Figure 10.10 certainly works, costs nothing, but involves more work and more time.

A simple turntable is another cost-effective and practical solution. It can be operated by a patient volunteer behind the screen (Figure 4.6) or automated as was done at Samsung (McMullin et al., 2015, figure 14). Versions of these automated movers can be built into walls for evaluating flat-panel TV audio systems, sound bars, on- and in-wall/ceiling loudspeakers, etc. (Olive, 2009; McMullin et al., 2015). SenseLab has a Rapid Speaker Spinner product

(https://forcetechnology.com/en/services/acoustics-noise-sound-quality/senselab-speaker-spinners [accessed September 2024]). There is an ambitious hobbyist version described in an audio forum, MatthewS (2023), so corporate funding is not any longer a requirement. Anyone performing frequent subjective evaluations is well advised to make such an investment. Reviewers take note.

10.3.5 Solutions for Multiple Listener Tests

Ideally, listen in a single room, using a single listener in one location. But there will be situations where this is not possible, or expedient. If there are multiple listeners, on successive evaluations it is necessary to rotate listeners through all of the listening positions so that they all get exposed to the variations in bass response in that particular room.

With multiple listeners it is difficult to avoid a "group" response. Some people don't mute verbal sounds signaling their feelings, and others may exhibit body language conveying the same message. This is especially true when anyone in the group is respected for his or her opinion. Such listeners sometimes deliberately reveal their opinions verbally or non-verbally, in which case the test is invalid. Enforce a "no discussion" rule between tests. Single-listener tests are much preferred.

Seats should have backs that rise no higher than the shoulders to avoid reflections and shadowing from reflections and multichannel surround loudspeakers.

10.3.6 Making the Test Blind and Double-Blind

The importance of blind listening was established in Section 10.1.1, but how, in practice, is it achieved? An acoustically transparent screen is a simple device for revealing truths, but it is essential to test the acoustical transmission loss of the material by performing a simple on-axis frequency response measurement, then place the material between the loudspeaker and microphone, and measure again. There will be a difference, but it should be less than 1 dB, even at the highest frequencies, as it is in a good grille cloth. Polyester double-knit is a commonly available possibility, and there are dedicated grille cloth fabrics in many fabric stores. However, care must be

taken to ensure that the fabric when tested and used is under slight tension to open the weave—it is this that is responsible for the acoustical transparency. Because fabrics with low acoustical loss tend to be very porous (easy to blow air through and to see glints of daylight through) they can reveal outlines and sometimes more of the loudspeakers being evaluated. Minimize the light behind the screen and emphasize the lighting levels on the audience side to alleviate this problem—directional lighting helps. Be sure to cover any LED power indicators and other lights on active loudspeakers.

Should you listen with the grilles on or off? It depends on your audience and the product. In reality, only audiophile enthusiasts are likely to want to, or can get away with, leaving the transducers exposed. In good products the differences should be vanishingly small, but this is not always so. Conscientious loudspeaker designers include the grille cloth and frame (often the bigger problem) into the design. In the real world, visual aesthetics count for something, and grille cloths are a deterrent to probing fingers.

All that preparation achieves the first level of blind testing—the listeners do not know what they are listening to. The second level, called double-blind, ensures that the experimenter cannot influence the results by structuring the presentations in a specific manner. This requires randomization of the important variables: program, loudspeaker identifiers, presentation sequence (unless it is a self-controlled single-listener test). All can be randomized in a computer controlled system. That was not possible in my early tests, so I created instruction sheets for experiment control to be followed by the operator, using random numbers to specify the sequences (random numbers are available on the internet).

If one listener in a group is permitted to do the switching, it is soon evident which loudspeakers are favored and which are not, so other listeners are biased, especially if that person has a “reputation”. An independent person controlling the experiment, following randomized sequence instructions, is the best solution for groups. Now a computer and automation can do it all.

10.3.7 Binaural Record/Replay Loudspeaker Comparisons

Many situations do not lend themselves to real-time, in situ listening tests. Binaural recordings reproduced at a later time provide an obvious alternative,

but many people have a problem accepting the idea that headphone reproduction is a satisfactory substitute for natural listening. The author also had doubts in the beginning, but the appearance of the Etymotic Research ER-1 insert earphones provided what seemed to be a solution to the largest problem: delivering a highly predictable signal to the listeners' eardrums. Toole (1991) describes the problems in detail, and shows results of early tests in which there was good agreement between real-time and binaural listening tests of some loudspeakers. The same basic process was used in experiments comparing results of listening tests done in different rooms (Sections 5.4 and 5.5 and Olive et al., 1995). The results of real-time and binaurally time-shifted tests yielded closely similar results in repeated tests and allowed for different kinds of tests to be performed, adding knowledge.

These tests employed a KEMAR head and torso simulator (an anthropometric mannequin) as a binaural microphone, and a problem was that much or all of the time the sound was perceived to be in-head localized. Playing the binaural recordings in the same room that they were recorded in significantly alleviated this, allowing listeners to see and hear the room before the recorded version took over. In such instances the sounds were often credibly externalized. However the illusion deteriorated when the listening was done in a different room and what the eyes saw did not match what the ears heard.

A more elaborate scheme called Binaural Room Scanning (BRS) involves measuring binaural room impulse responses (BRIR) at each of many angular orientations of the mannequin in the test environment. For playback the music signal is convolved with the BRIR filters as it is reproduced through headphones attached to a low-latency head-tracking system. By this means, as the listener's head rotates, the sounds at the ears are modified so as to maintain a stable spatial relationship with the real room. The result is a much more persuasive externalization of the sound and the associated spatial cues. Olive et al., 2007; Welti and Olive, 2013; Gedemer and Welti, 2013; and Gedemer, 2015, 2016 describe such experiments in domestic rooms, cars and cinemas. The Smyth Realizer is a commercial implementation (https://smyth-research.com).

A literature search will reveal many more examples, because binaural techniques are now accepted in audio and acoustical research. A fact not recognized by many skeptics is that this requires a precise knowledge of the sounds that reach the listeners' eardrums. A carefully calibrated system is required and headphones or, preferably, insert earphones are chosen for the ability to deliver predictable sounds to the eardrums. HRTF mismatches to individual listeners are inevitable, but it is important that to remember that these tests are *comparative* multiple-loudspeaker evaluations, not *absolute* evaluations of sound quality, which means that constant errors of any origin, including recordings, are significantly normalized.

10.3.8 Equal-Loudness Testing at a Known Loudness Level

In any listening comparison louder sounds are perceptually recognizable which is an advantage for persons looking for a reason to express an opinion. Other things being equal, that sound would also have more perceived bass (the equal-loudness contours, Section 9.4). It is a well-known sales strategy. If the frequency responses of the devices being compared are identical, as in most electronic devices, loudness balancing is a task easily accomplished with a simple signal like a pure tone and a voltmeter. Loudspeakers radiate distinctive three-dimensional sound fields into a reflective listening space, meaning that it is probably impossible to achieve perfect loudness equality for all possible elements (e.g. transient and sustained) of a musical program.

There has been a long quest for a perfect "loudness" meter. Some of the offerings have been exceptionally complicated, expensive, and cumbersome to use, requiring narrow-band spectral analysis and computer-based loudness-summing software. The topic is discussed in Section 10.4.

The practical problem is that if the frequency responses of the loudspeakers (or headphones) being compared are very different a loudness balance achieved with one kind of signal would not apply to a signal with a different spectrum. Fortunately, as loudspeakers have improved and are now more similar, the problem has lessened, although it has not disappeared. In all these cases, if in doubt, turn the instruments off and listen; a subjective test is the final authority—it is used to test the validity of measurements.

Once we have achieved “equal loudness” among the loudspeakers being compared, the next step is to decide on a playback sound level, and then maintain it for all programs for all presentations.

Spectral balance is affected by playback sound levels, and so are several other significant perceptual dimensions: fullness, spaciousness, softness, nearness, and the audibility of distortions and extraneous sounds, according to Gabrielsson et al. (1991). Higher sound levels permit more of the lower-level sounds to be heard, notably those related to spatial perceptions. Therefore, in repeated sessions of the same program material, the sound levels must be the same. Allowing listeners to find their own “comfortable” playback level for each listening session may be courteous, but it is bad science.

Of course, background noises can mask lower-level sounds. This is a special concern in automotive audio, where perceptual effects related to timbre and space change dramatically in going from parked to highway speeds.

Recent investigations found that conventional TV dialog is satisfying at an average level of 58 dBA in a typical domestic situation, but at 65 dBA in a home theater setting. In home theaters, different listeners expressed preferences ranging from about 55 to 75 dBA (Benjamin, 2004). Using a variety of source material, Dash et al. (2012) concluded that a median level of 62 dB SPLL was preferred (using ITU-R BS.1770 weighting. It can be approximated by B or C weighting). The preferences ranged from 52 to 71 dB SPLL. These two studies are in basic agreement, meaning that satisfying everyone in an audience may be impossible. Frequency weighting is explained in Section 10.4.

These are not high sound levels compared to those used by many audiophiles or movie viewers in foreground listening mode. Cinemas are calibrated to a 0 dB reference of 85 dB (C-weighted), which, allowing for 20 dB of headroom in the soundtrack, permits peak sound levels of 105 dB—for each of the front channels, and every channel in an Atmos system. In cinemas this can be *very* loud. Low-Frequency Effects (LFE) channel output can be even higher, so it is not surprising that sustained levels of this combined sound are unpleasant to some. Patrons have walked out of movies, causing some cinema operators to turn the volume down. This compromises the art and

degrades dialog intelligibility. As discussed previously, it is a serious industry problem.

This same calibration process is used in many home theater installations. Many (most?) listeners find that movies played at "0" volume setting to be unpleasantly loud. Home theater designers and installers in my CEDIA classes routinely reported that their customers play movies at −10 dB or lower. That is a power reduction of a factor of 10—a significant cost factor when equipping such a theater. Even a reduction of 3 dB is a factor of 2 in power.

Recommendation ITU-R BS.1116–2 (2014) chose to standardize the *combined* playback level from some number of active channels at 85 dBA, allowing 18 dB before digital clipping. Bech and Zacharov (2006) comment that they find this calibration level to be excessively high by 5–10 dB for subjects listening to typical program material.

So, the recommended calibrations for playback sound levels are frequently found to be excessively high, as judged by moviegoers and in private home theaters. The point of this discussion is that the choice of a playback sound level for product evaluation is a serious matter. It affects what can be heard, and how it is heard (see the equal-loudness contours in Section 9.4). Ideally, comparative listening should be done at a sound level that matters to you, or to the audience for your opinions. Olive et al. (2013) used program at an average level of 82 dBC, slow. This was for an audience of serious listeners doing foreground listening of a demanding kind. Most tests do not reveal their listening levels, so we never really can be certain what the results mean.

Music and movies are often mixed while listening at high sound levels. When I have walked into these sessions I sometimes found myself quickly putting my fingers in my ears and retreating to the back of the room. I value my hearing. Hearing loss is a known occupational hazard in professional audio. Naturally, when these recordings are played at "civilized" sound levels the bass is perceptually attenuated because of the equal-loudness contours. Unless this is compensated for in the mastering process, a bit of bass boost may be required to restore natural balance. This is why I tediously recommend having easy access to, and using, a bass and/or subwoofer level control.

10.3.9 Choosing Program Material

The ability to hear differences in sound quality is greatly influenced by the choice of program material. The audibility of resonances is affected by the repetitive nature of the signal, including reflections and reverberation in the recording and in the playback environment (Toole and Olive, 1988). The poor frequency response of some well-known microphones is enough to disqualify recordings made with them, whatever the merits of the program itself (Olive and Toole, 1989; Olive, 1990). Some of the historically popular large diaphragm units have frequency responses at high frequencies that would be unacceptable in inexpensive tweeters.

Olive (1994) shows how, in training listeners to recognize resonances in loudspeakers, it is possible to identify the programs that are revealing of differences and those that are merely entertaining. Figure 10.13 displays the programs that Olive used in those tests, and the effectiveness of those programs in allowing listeners to correctly recognize spectral errors, bass and treble tilts and mid-frequency boosts and cuts, introduced into the playback. The important information is not specific to the recordings themselves, because more modern recordings should perform similarly, but to the nature of the program itself. The more complex and spectrally balanced the program, the more clearly the spectral errors or differences were revealed to listeners. It is surprising that some types of program that are highly pleasurable, or important to a genre—such as speech in movies and broadcasting—are simply not very revealing, and therefore not very demanding of, spectral accuracy.

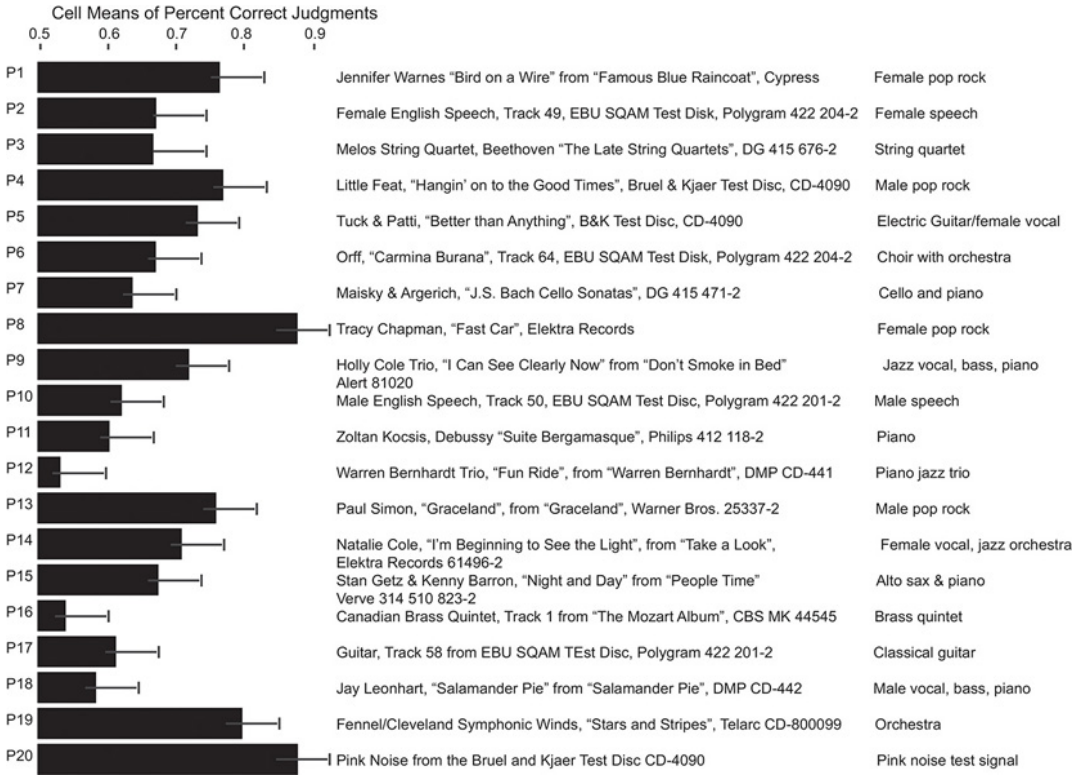


FIGURE 10.13 *The ability of different kinds of program to reveal spectral errors in reproduced sound. From data in Olive (1994).*

Figure 10.14 shows that an important requirement for good program material is a long-term spectrum that well represents most of the audible frequencies, presumably giving listeners more opportunity to judge. What they were judging in these tests was the ability to hear narrow and broadband spectral variations in loudspeakers—resonances—arguably the problems most disruptive of sound quality/timbre.

The shapes of the curves tell us that listeners perform best when the program has strong energy at all frequencies, the better programs showing the flattest curves. The inversion of the bass contents of Groups 3 and 4 breaks the trend: Group 4 had more low-frequency content than Group 3, indicating that there is more to this than long-term spectral content. Vocal/instrumental composition is another factor, some of which can be seen in the program listings in Figure 10.13. And, finally, there is the inescapable factor of the "circle of confusion"—we don't know what spectral irregularities are incorporated into the recordings. More specific details are in Olive (1994).

Group 1: pink noise; female pop rock; symphonic wind orchestra; male pop rock; female pop rock.

Group 2: male pop rock; female vocal with jazz orchestra; alto sax and piano jazz; electric guitar and female pop vocal; female vocal, bass, piano jazz.

Group 3: male speech; female speech; string quartet; choir and orchestra; cello and piano.

Group 4: solo piano; piano jazz trio; brass quintet; classical guitar; male vocal with piano and bass jazz.

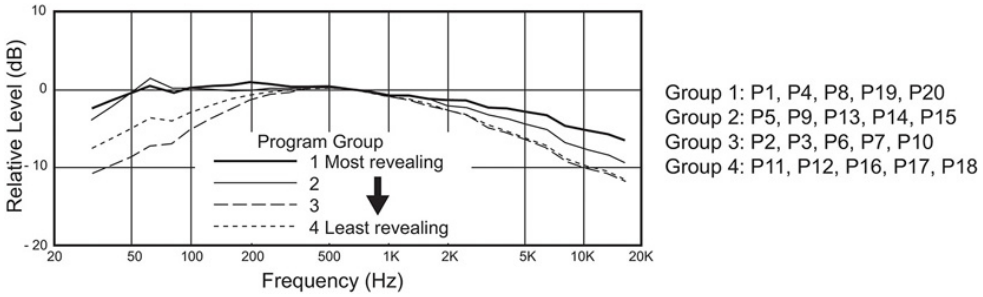


FIGURE 10.14 *Long-term 1/3-octave spectral averages of programs ranked according to their abilities to reveal differences in loudspeakers. From Olive (1994).*

Clearly there are no hard rules but, in general, complex productions with broadband, relatively constant spectra aid listeners in finding problems. The notion that "acoustical", especially classical, music has inherent superiority was not supported. Solo instruments and voices appear not to be very helpful, however comfortable we may feel with them. A variety of programs should be used. Amusingly, a program that is good for demonstrations (e.g. selling) can be different from that which is needed for evaluation (e.g. buying).

The efficiency and effectiveness of listening tests improves as listeners learn which aspects of performance to listen for during different portions of familiar programs. Select a number of short excerpts from music known to be revealing, and use them repeatedly in prerecorded randomized sequences. They can be edited and assembled on computers with pleasing segues to complete loops. It is not entertainment, but the selections must have some musical merit or they will be annoying. After many, many repetitions the

music becomes just a test signal, but because of the familiarity, it is extremely informative. Because the circle of confusion is a reality, it is essential to use several selections from different sources, recording engineers, labels, etc.

By paying attention, it will soon become clear which selections are helpful, and which are just using up time. Ask your listeners what they use to form their opinions. It has been known for listeners to have a very narrow focus—like using only the kick drum, or ignoring anything classical, or ignoring anything popular, or listening only to voices. Such listeners should not be invited back.

It is also not uncommon for the broadband individualities of certain programs to be noticed and accepted as "normal" program variations. For example, one pop/rock excerpt that was very revealing of spectral features was expected to have slightly too much low bass, others might be known to be slightly bright or dull, the kinds of spectral effects one might correct with a tone control if one was fussy. If a spectral feature is heard in many loudspeakers in multiple-loudspeaker comparisons, it is assumed to be part of the program, not something added by the loudspeaker. This relates to the circle of confusion (Figure 1.8) in which such variations, especially in bass level, are common. The important conclusion is that, though there are broadband spectral balance differences among recordings, the recordings are still able to reveal resonances—technical flaws—in loudspeakers. It would seem that the sound quality aspect of "translation" is fundamentally communicated, although broadband spectral balance may need adjustment at the listening end.

10.3.10 Electrical and Electronic Decisions

It is usually safe to assume that electronics are "transparent" audio devices, but there are exceptions and special circumstances that warrant our attention to avoid unwanted influences in the results of listening tests. These influences range from subtle changes in the spectrum of sounds to momentary distortions that can be alarming. These are electrical engineering issues, but they routinely get elevated to different planes of thought by subjective reviewers, even to the point of considering certain problems to be advantages.

Impedance: 8 ohms. This is the kind of specification one sees for loudspeakers. The number may be different, but whatever it is, it is an approximation. For a few—very, very few—loudspeakers it is a good approximation, but for the vast majority it is a poor description of reality. Figure 10.15 shows a typical loudspeaker impedance. It varies substantially with frequency and it matches the rated impedance at four frequencies. The variations are normally of no concern unless they drop to very low values, and tradition requires that a number be specified.

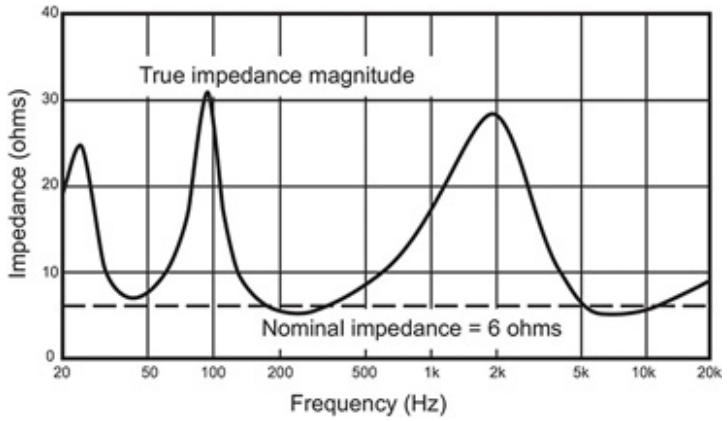


FIGURE 10.15 *An impedance curve for a loudspeaker compared to the nominal impedance rating chosen by the manufacturer for it. The minimum impedance should also be specified.*

Power amplifiers should be constant-voltage sources—the output voltage should be independent of the load impedance. There are limits, though, relating mainly to the current available from the power supply if the load impedance is especially low. Occasionally, an unfortunate interaction between amplifier and loudspeaker provokes current or voltage limiting or the activation of a protection circuit, all causing momentary audible effects during crescendos. Most power amplifiers, including many receivers (check the specification for minimum load impedance) can drive a minimum impedance of 4 ohms comfortably, and some can drive 2-ohm minima. Unfortunately, it is not always specified.

The loudspeaker shown in Figure 10.15 should be easy to drive, but some products have impedance minima that drop to lower values. The minimum

impedance is an important number.

There have been some examples of high-end loudspeakers having impedances that dipped to one ohm or less. These are examples of incompetent loudspeaker design. However, recognizing a market, some amplifier designers responded with monster monoblock “arc-welder” products that can drive these problem loudspeakers, but they are more than is needed for well-designed loudspeakers. It was amusing at the time to read that these imperfect loudspeakers “revealed” differences between power amplifiers, as if it were a virtue. They were the *cause* of the differences. But those were subjective reviewers not having the benefit of measurements. Before conducting listening tests it is beneficial to know the minimum loudspeaker impedance and the capabilities of the power amplifier driving them. The ability to drive 2 ohms is not uncommon and should be adequate.

It was stated earlier that the dramatic impedance variations were normally of no importance. This is because most power amplifiers are approximate constant-voltage sources, meaning that they have very low output impedances (also called internal impedance). The load current can vary, but the output voltage remains stable. But what if this is not the case?

To explain the situation, we need to resort to some elementary electrical engineering. Figure 10.16 shows simple schematic diagrams of the output of a power amplifier, driving a loudspeaker through speaker wire.

Figure 10.16 (a) and (b) show that the output impedance of the power amplifier and the resistance of the loudspeaker wire are components in a voltage divider circuit. When combined with the frequency-dependent impedance of the loudspeaker it means that the “flat” frequency response voltage at location “A-A” inside the power amplifier acquires a shape following that of the impedance curve at location “B-B”. Because this is the voltage driving the loudspeaker, the overall performance of the loudspeaker, i.e. all its frequency response curves, are modified by this amount. Different loudspeakers have different impedance curves; some are strikingly variable, others change little.

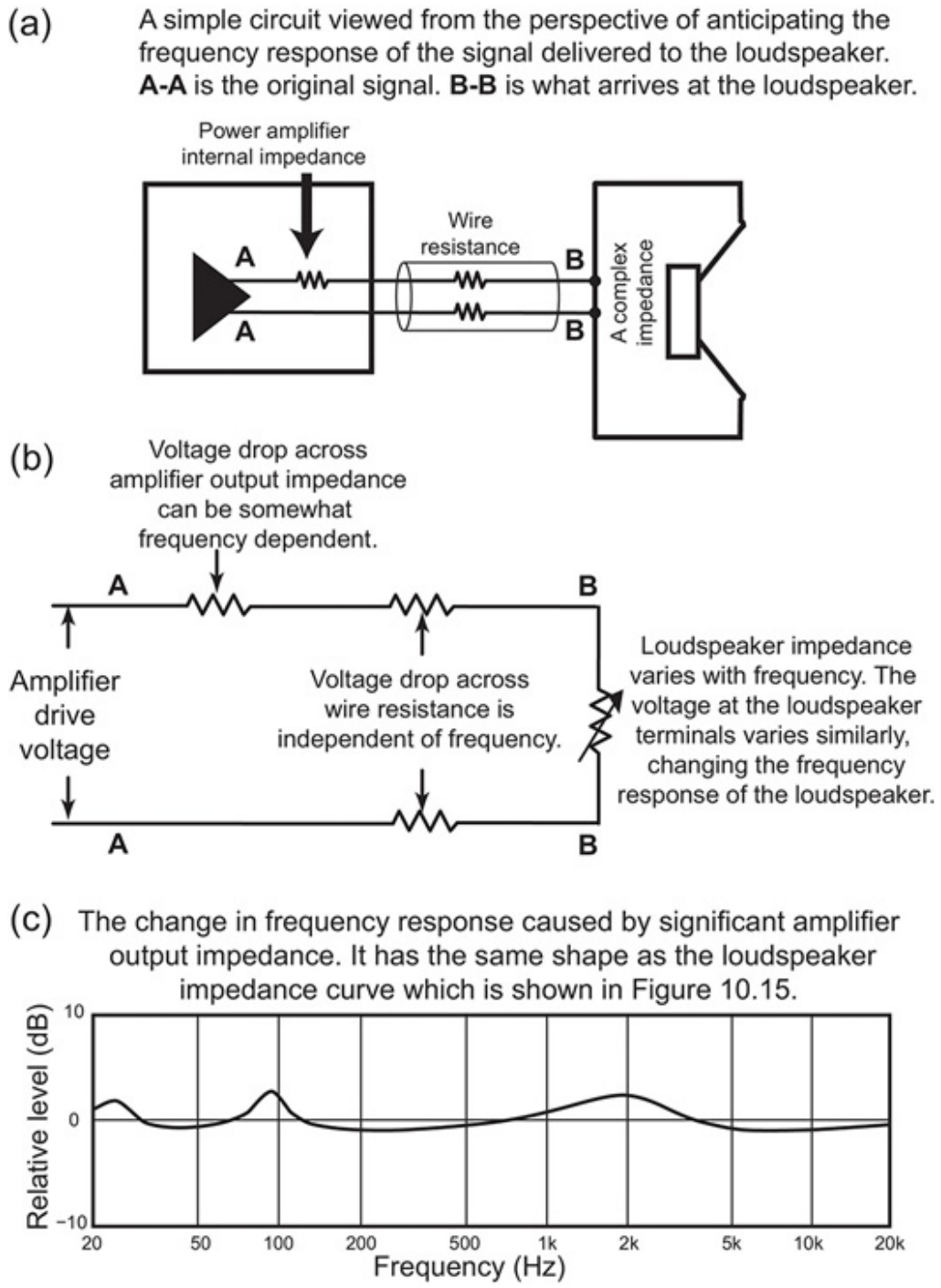


FIGURE 10.16 *Schematic diagrams showing (a) and (b) the electrical circuit explaining how amplifier and wire resistances cause variations in loudspeaker frequency response, and (c) how loudspeaker frequency response can be altered by the high output impedance of a tube power amplifier.*

To give this diagram more meaning, let us review the actual numbers that apply to the illustrated “resistors”. The amplifier data are from numerous manufacturer’s specifications and *Stereophile* review measurements (thank you John Atkinson for measuring useful data)—an extensive, but not an exhaustive survey.

Output impedance, also called *internal impedance*, is rarely specified but it sometimes appears as a standardized calculation called *damping factor*. Some manufacturers with highly recognizable brand names disclose neither specification. A low impedance or a high damping factor are desirable specifications.

Damping Factor (DF) = 8 □ output impedance

- Output impedance: Class AB solid-state power amplifiers: typically, 0.01 to 0.2 ohm (DF 800 to 40) with outliers to 0.5 ohm (DF 16). Class D: 0.01 to 0.11 ohm (DF 800 to 80).
- Output impedance: tube power amplifiers: typically, 0.3 to 3.3 ohms (DF 27 to 2.5), occasionally higher, max observed 10+ ohms (DF > 0.8).
- Speaker wire resistance: 10 ft (3 m) of 12 gauge, both conductors, 0.032 ohm.
- Speaker wire resistance: 10 ft (3 m) of 14 gauge, both conductors, 0.052 ohm.
- Speaker wire resistance: 10 ft (3 m) of 16 gauge, both conductors, 0.082 ohm.
- Speaker wire resistance: 10 ft (3 m) of 18 gauge, both conductors, 0.148 ohm.
- Loudspeaker impedance: typically, 3 ohms to 30 ohms varying with frequency.

The amount of the change in frequency response depends on the total voltage drop across the combined amplifier output impedance and wire resistance, meaning that minimizing both of these is desirable. Wire resistance is negligible for any reasonable gauge and distance.

To reviewers these are moderately discomfiting numbers, because the inevitable conclusion is that many tube power amplifiers cannot allow loudspeakers to perform as they were designed. Of course these tube amplifiers sound "different" from solid-state amplifiers, because they modify the loudspeakers. Reviewers handle the issue in various ways. Some ignore it, others have danced around the matter concluding that it is just one more uncertainty in sound reproduction. Rarely is it acknowledged to be what it is. One thing is certain, for meaningful comparative loudspeaker evaluations solid state power amplifiers are generally advantageous but try to verify the output impedance/damping factor specification.

One well-known high-end loudspeaker specified that it should be used with wire having resistance less than 0.2 ohms. This conscientious advice is admirable, but it means that the restriction is violated as soon as any tube amplifier is connected, no matter what wire is used. It is sobering to think that 10 ft (3 m) of the much despised 18-gauge lamp cord has 0.148 ohms resistance—much less than the output impedance of typical tube amplifiers.

Minimizing wire resistance is easy: use large wire, having low gauge numbers—see Table 10.1—or, better yet, use less wire. If there is a risk of radio-frequency signal pickup, it is important to know that unshielded wires act as antennas. A great deal of mystique has evolved around loudspeaker wires, attempting to elevate this simple device to impossible heights of importance. Notions that they behave as transmission lines persist, but Greiner (1980) offers persuasive arguments that this is unrealistic. There are other beliefs, some of which are impossible (e.g. directional wires), or irrelevant (e.g. skin effect, which is significant only at frequencies much above the audio frequency band). The little wire stands are especially quaint.

TABLE 10.1 Resistances per unit length of two-conductor stranded copper wire, stated for BOTH wires in the circuit—so just measure how long the two-conductor wire is, and multiply by these numbers. For reference, common lamp cord is typically 18 gauge. If you do not see a gauge rating for a loudspeaker wire, be very suspicious. Some exotic cables use small wire for seriously mistaken reasons. ⏎

AVG Wire gauge	Resistance per ft (ohms) BOTH conductors	Resistance per m (ohms) BOTH conductors
10	0.0020	0.0067
12	0.0032	0.0106
14	0.0052	0.0169
16	0.0082	0.0268
18	0.0148	0.0483

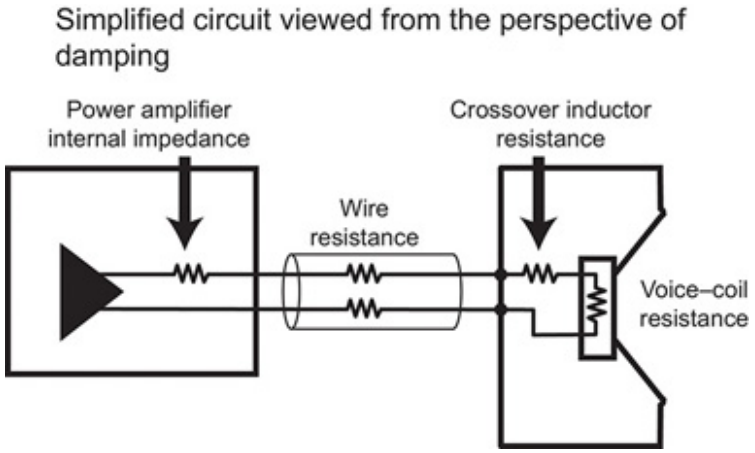


FIGURE 10.17 *A schematic diagram of a power amplifier connected to a passive loudspeaker, as in Figure 10.16(a), but this time showing the resistances of the woofer inductor and voice coil within the interior of the enclosure that complete the series electrical circuit. This complete loop of components determines that electrical damping of the woofer diaphragm.* ⏎

At prices that can exceed $20,000 for a pair of 8-foot (2.4 m) loudspeaker wires, one expects a lot, but believing is hearing as Section 10.1 explains. Wire is a good product for the industry: totally reliable, inexpensive to manufacture, and highly profitable.

One of the universal compliments reviewers attach to almost any audio product they like is that it results in “tighter bass”. Damping unwanted motion

of a loudspeaker diaphragm is undoubtedly a good thing, and the term damping factor implies that it has some relevance.

In 1975 I wrote an article for AudioScene Canada magazine called "Damping, Damping Factor and Damn Nonsense". I still like the title because is a succinct statement of reality.

Figure 10.17 shows the complete circuit involved in the electrical damping of loudspeakers—it does not mysteriously stop at the loudspeaker terminals. The power amplifier damping factor (determined by output impedance) is not alone in deciding system damping. Current must flow through components and devices inside the enclosure. After flowing through the wire, it typically passes through an inductor, part of the low-pass filter ahead of the woofer in a passive system. Then, inside the woofer there is the voice coil. The inductor resistance is commonly around 0.5 ohm, and the voice coil resistance can have different values but is commonly around 6 ohms. So, let us examine all of the resistances in the circuit to arrive at the following progression of damping factor changes:

Amplifier internal impedance: 0.01 ohm	DF = 800
Add wire resistance, 10 ft of 10 gauge, both conductors: 0.02 ohm	DF = 266
Add crossover inductor resistance: 0.5 ohm (typical)	DF = 15
Add voice-coil resistance: 6 ohms (typical)	DF = 1.2

Obviously the resistances inside the loudspeaker are the dominant factors. Even eliminating the inductor and driving the woofer directly changes things only slightly. The article (Toole, 1975) shows oscilloscope photographs of tone bursts of various frequencies and durations while the damping factor of the amplifier was varied from 0.5 to 200. At damping factors above about 20 (internal impedance less than 0.4 ohms), no change was visible in any of the transient signals and changes in frequency response were very much less than 1 dB, and then only over a narrow frequency range. On music no change in sound quality could be discerned, including attentive listening for "tightness". This was not a research project. It was an exercise to satisfy my curiosity, and

it was satisfied in that my instinct that damping factor was a misnomer was correct. Damping is important, but amplifier damping factor is but one element, and not the dominant one, in achieving it.

Adequately sized hardware store speaker wire (12 gauge for long runs, 14 gauge for shorter runs) and popularly priced audio frequency analog interconnects are transparent—in spite of colorful marketing stories. Up-market interconnects sometimes have connectors that make more aggressive electrical contact, but beyond that one enters the realm of with what I call audio "jewelry". Digital communication and interfacing is a different and increasingly complex matter—and outside my comfort zone to discuss. I know from personal experience that HDMI handshakes are a common audio/video problem.

10.3.11 Listener Choice and Training

Not all of us are good listeners, just as not all of us can dance or sing well. Some of it is aptitude, but, as has been made clear in earlier sections, if a listener suffers from significant hearing loss that person's opinions will differ from those of the majority. They will be less reliable, exhibiting higher than normal variations in repeated exposures to the same sound, and personal biases depending on the nature of the hearing abnormality. Age is a factor in this, as well as exposure to loud sounds; see Section 9.2.2. Audiometric hearing thresholds within 15 to 20 dB of normal would seem to be a reasonable criterion from this perspective. If hearing is defective, no other qualifications matter.

If one is establishing a population of listeners from which to draw over the long term, it will be necessary to monitor their decisions, looking for those who (a) exhibit small variations in repeated judgments and (b) differentiate their opinions of products by using a large numerical range of ratings (Toole, 1985; Olive, 2001, 2003). An interesting side note to this is that the interests, experience, and, indeed, occupations of listeners are factors.

Musicians have long been assumed to have superior abilities to judge sound quality. Certainly, they know music, and they tend to be able to articulate opinions about sound. But, what about the opinions themselves? Does living "in the band" develop an ability to judge sound from the audience's

perspective? Does understanding the theory and structure of music and how it should be played enable a superior analysis of sound quality? When put to the test, Gabrielsson and Sjögren (1979) found that the listeners who were the most reliable and also the most observant of differences between test sounds were persons he identified as hi-fi enthusiasts, a population that also included some musicians. The worst were those who had no hi-fi interests. In the middle, were musicians who were not hi-fi oriented. This corresponds with the author's own observations over many years.

Perhaps the most detailed analysis of listener abilities was in an EIA-J (Electronic Industries Association of Japan) document from 1979 that is shown in Figure 10.18. It was the result of several years of observation and a very large number of listening tests (personal communication from a committee member). It would have been useful to see performance statistics, if they existed.

	Categories of listeners/audio	Sound analytical ability	Knowledge of reproduced sound	Attitude towards reproduced sound	Application objective	Reliability
1	General public	mediocre	mediocre	biased	various general investigations	
2	General public interested in audio equipment	mediocre	mediocre	biased	various general investigations	
3	General public strongly interested in audio equipment	fairly	some	strong self-assertion tendency	semi-specialized studies and measurements by specific group	low

	Categories of listeners/audio	Sound analytical ability	Knowledge of reproduced sound	Attitude towards reproduced sound	Application objective	Reliability
4	Audio equipment engineers	high	sufficient	most correct	specialized studies and high precision measurements	high
5	Experienced acoustic experts	adequately high	excellent	correct	various studies and high precision measurements	high
6	Musicians, including students	adequately high	some	over-rigorous or non-interest	valuable opinions but unsuited for measurement	
7	General public strongly interested in music	high	some	mostly uninterested	unsuited for study & measurements (opinions are valuable)	
8	General public interested in music	somewhat high	mediocre	biased	various studies as representative of the general public	
9	General public strongly interested in recorded music	fairly high	fair	roughly correct	various studies and measurements	fairly high

FIGURE 10.18 *A chart describing several aspects of listener capabilities and their suitability for different kinds of subjective audio evaluations. From EIA-J (1979).* ⏎

Olive (2003) analyzed the opinions of 268 listeners who participated in evaluations of the same four loudspeakers. Twelve of those listeners had been selected and trained and had participated in numerous double-blind listening tests. The others were visitors to the Harman International research lab, as a part of dealer or distributor training or in promotional tours. The results shown earlier in Figure 10.9 reveal that the trained, veteran listeners distinguished themselves by having the highest performance rating by far, but obviously years of experience selling and listening on store floors had a positive effect on the retail sales personnel.

Olive (1994) describes the listener training process, which is also a part of the selection process; those lacking the aptitude do not improve beyond a moderate level. It teaches listeners to recognize and describe resonances, the most common defect in loudspeakers. It does not train them to recognize any specific "voicing" as some have suspected. Instead, by a process of eliminating timbre-modifying resonances, the preferences gravitate to "neutral" loudspeakers, through which listeners get to hear the unmodified recordings.

10.3.12 Cultural and National Biases

It is reasonable to think that persons from different countries or cultures might prefer certain spectral balances. It is also common to think that people of different ages, having grown up with different music might have preferences that are not shared by young people. Harman is a large international company, with audio business in all product categories, consumer, professional, and automotive so it is important to know if such biases exist, and if they do, to be able to design products that appeal to specific clienteles. Over the years several investigations have shown that, if such biases exist at all, they are not revealed in blind testing.

Sean Olive has conducted numerous headphone sound quality evaluations with listeners from Europe, Canada, Japan and China and has yet to find

evidence of strong cultural bias. See Olive et al. (2014) and Chapter 15. Similar results have been found for loudspeakers and car audio. The same is true with respect to age, except, as shown in Figure 10.7, older listeners tend to exhibit greater variation in their opinions because of degraded hearing.

10.4 LOUDNESS AND SOUND LEVEL MEASUREMENTS

Listening tests require equal loudness among the compared loudspeakers in order for the comparisons to be fair. All tests should be conducted at a sensible program sound level that needs to be replicated in successive tests. Both need measurements.

There has been a long quest for a perfect "loudness" meter that yields a single number perfectly representing the subjective loudness effect of any sound, as heard in any environment. The practical problem is that music, movies and mixed broadcast programs are not stable. The sounds continuously change in bandwidth, spectrum, level and transient/temporal structure. If we add the psychoacoustic complications of loudspeakers with different frequency responses and directivities located in different rooms, at different incident angles relative to listeners, at different distances, the situation becomes much more complicated. It is important to examine what matters and what does not matter. Even though the perfect solution may not exist, we do have measures that permit us to achieve our practical needs.

Historically, the first simple approach to a single number loudness rating derived from the equal-loudness contours shown in Figure 9.5(a). The A-weighting curve shown in Figure 10.19, for example, resembles the shape of an inverted 40-phon loudness contour, and thereby was thought to be appropriate for measuring low-level sounds. B and C weighting curves were thought to be more representative of progressively higher sound levels. It didn't work as intended and, over the years, C-weighting came to be used as an approximation to flat, but with some discrimination against very low and very high frequencies. A-weighting has become widely accepted as a general-purpose measure where low frequencies are not an issue, including assessment of hearing damage risk for occupational hearing conservation programs. B-

weighting has been all but forgotten and is no longer a standard feature in sound-level meters.

It really should not have been surprising when schemes based on equal-loudness contours should be found deficient. After all they were derived from comparisons between pairs of pure tones in an anechoic environment, not a continuous broadband program with continuously varying temporal and spectral content in a semi-reflective normal room.

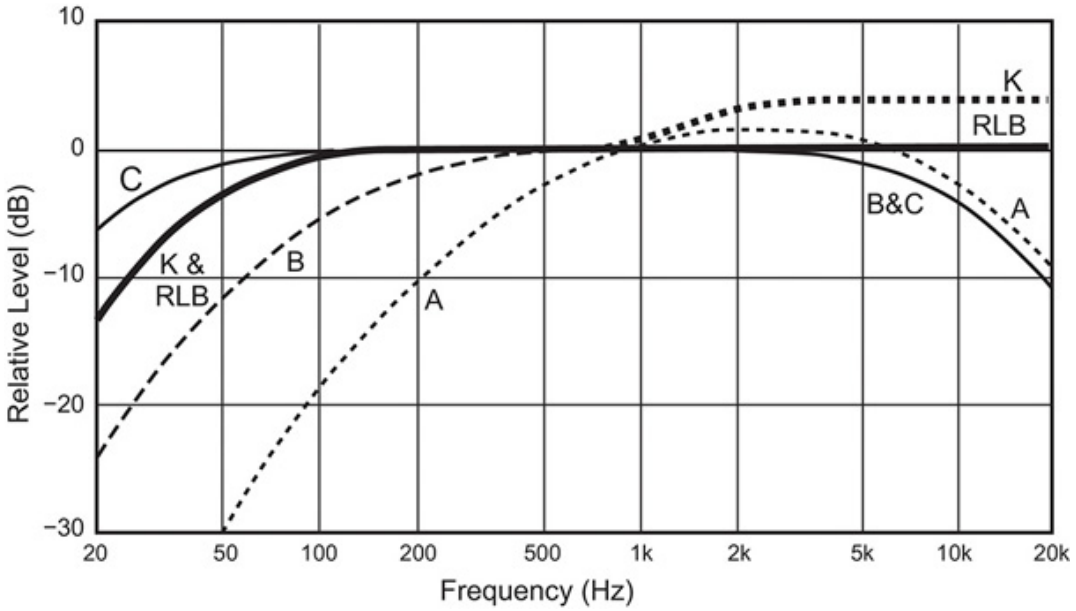


FIGURE 10.19 *Frequency-weighting curves that are available in sound-level meters (A, B and C) and the RLB curve proposed by Soulodre and Norcross (2003) for providing a single-number measure of the perceived loudness of typical audio signals. K weighting is the RLB curve with a high-frequency boost as used in the ITU-R BS.1770–5 (2023) loudness recommendation.* ⏎

Substantial research over many years resulted in much more elaborate methods for computing loudness as described in Fastl and Zwicker (2007) and Moore et al. (1997). They begin with narrow-band spectral analysis and loudness-summing procedures based on psychoacoustic models of some the frequency, amplitude and masking properties of the hearing process. The complexity of such models appeals to some, but the supporting data are limited and include large statistical variations in subjective responses. As

stated by Moore et al. (1997) in section 3, describing limitations of their model, “the predictions of the model should not be expected to be accurate for individual listeners or even small groups of listeners”. That paper gives an excellent overview of the variables considered in these models, showing just how complex the process is. The real-world problem is that so many variables insist on varying.

One of the main driving reasons for loudness evaluations is to maintain relatively stable loudness experiences from recordings of many kinds from many sources. In broadcast programming there is a long history of devices created for monitoring program levels—the row of meters on the meter bridge of a recording or broadcast console for example. One critical need is purely technical; to avoid overloading something in the signal path for which a peak signal level indicator is needed. The other is to prevent listeners from being presented with inappropriately loud or quiet passages for which a longer-term averaging meter is needed.

Given the complexities and uncertainties of the modeling options, it was encouraging when research suggested that a much simpler solution might be sufficient (Soulodre, 2004; Soulodre and Norcross, 2003). It is an easy to implement high-pass characteristic somewhere between B and C weighting but with no high-frequency roll-off. Figure 10.19 shows the standard A, B, and C weighting curves, along with their new proposal, the RLB (Revised Low-frequency B) curve.

The RLB curve provides the basis for “K” weighing used in ITU R-REC-BS.1770-5 (2023), which adds a 4 dB boost above about 1 kHz to accommodate the acoustical gain added by a spherical head. This international standard is aimed the radio and TV broadcast industries and embraces reproduction in mono, stereo and multichannel formats. Measurements using this standard are called LKFS, meaning Loudness K-weighted Full Scale. It is extensively used in the professional audio recording and streaming industries. Users set a reference level they wish to maintain, and monitor progress using LUFS, Loudness Units Full Scale, to follow real-time loudness levels compared to that reference. It is available as a plug-in for digital audio workstations, stand-alone instruments, and smart-phone apps.

In practical terms for setting up listening tests there are options. Following international standards, one can employ LKFS measures. If one already has a sound level meter with C weighting, Figure 10.19 shows that this weighting responds slightly more to very low frequencies and a lot less to high frequencies. Is it an important difference? Numerically, yes. Perceptually, it would seem so, but the following information raises doubts.

Before the current ITU standard was created, a group in Europe seriously investigated loudness calibrations for multichannel systems, something also relevant to setting up listening tests. They were extremely thorough, using a number of very different frequency weightings and bandwidths, including the complex calculated Zwicker and Moore models. Perspective on the series of experiments can be seen in Zacharov (1998) and Bech (1998). More details are reported in a series of five AES papers.

- Part I (Suokuisma et al., 1998) describes the test signals that are used.
- Part II (Zacharov et al., 1998) considers the influence of signals and loudspeaker placement.
- Part III (Bech and Zacharov, 1999) consider the effects of loudspeaker directivity and reproduction bandwidth.
- Part IV (Zacharov and Bech, 2000) investigates the correlation between measurements and subjective judgments.
- Part V (Bech and Cutmore, 2000) looks at reproduction level, the room, step size and symmetry.

Consistent throughout the experiments was a comparison of the following very different test signals (my descriptions):

- 700 Hz narrow-band noise
- 250–500 Hz bandpass-filtered noise
- 500 Hz–2 kHz bandpass-filtered noise (as used in many consumer products)
- Pink noise (as used in cinema calibrations)

- B-weighted noise
- Four broadband noise spectra based on Zwicker and Moore loudness models

Listeners matched loudness between loudspeakers in different locations using these signals and their adjusted levels were compared to measured levels. Altogether it was an impressive undertaking. There are many details in the results, but some substantive conclusions were:

- In *symmetrically arranged systems* with identical forward-firing loudspeakers, the choice of test signal for subjective adjustments of loudness was *not* a factor. This was remarkable given the huge variations in the options. Listeners adjusted loudness to match that of the forward-firing center channel. Repeated adjustments for the L and R and surround channels had standard deviations in the range 0.4 to 0.6 dB. (Part II).
- In *asymmetrical setups* loudspeaker distance was a strong factor. Only the narrow-band noise signal deviated from the pattern of the other test signals. Listeners were able to match loudness levels among the channels with standard deviations of under 0.4 dB. (Part II).
- A comparison of two forward-firing loudspeakers with significantly different frequency-dependent directivities (Genelec 1030A and B&O BL6000) and a Quad ESL63 dipole yielded no significant influence on the level adjustments for the front channels, and no significant influence when used as surround channels when the prime axis of the Quad was aimed at the listeners. The technical and subjectively evaluated level differences were similar.
- Pulling together all of the above data and more, correlations between subjective and objective measures were performed to find the combination of test signals and measurements that best predict subjective responses. (Part IV) The results indicate the following as being optimal (my comments in italics):

- Test signal: constant broadband specific loudness signal according to the Zwicker free-field model. *Broadband signal is good, but B and C weighting narrow the band and 500 Hz high pass filtering removes almost half of the spectrum.*
- Metric: Moore or Zwicker (diffuse or free field) loudness, or B- or C-weighted sound pressure levels. *Interesting that the most complex and the simplest metrics are competitive.*
- 500 Hz high-pass filtering any of the test signals improved the correlations slightly. *This eliminates all bass, and room resonance issues—was that a factor?*

It is important to know that even the less successful test signals and metrics were not in error by an amount that is likely to degrade entertainment value. This is especially so when the numerous variables in the production of programs are included. The spread of average values was typically in the range of 1 to 2 dB.

It is clear that the perception of comparative loudness levels is a very durable phenomenon, showing considerable consistency, which is good because we need to do it. Listening tests have fewer variables than were explored in the above experiments. All the sounds will come from the front at the same distance, and if positional substitution is used, from the same location. The main remaining variable is likely to be loudspeaker directional patterns, and even then the effect is not likely to be great. Given the range of test signals employed in the experiments just described, it is not likely that using the ITU R-REC-BS.1770-5 (2023) metric would have changed the results one way or another. It has prestige, "brand recognition", but in the end it is just another weighting curve.

It is important to remember that in evaluating all these loudness measurement schemes, the reference has been a purely subjective judgment. So, in the absence of any instrumentation, an unbiased subjective comparison of loudness is irrefutably correct but is likely to be somewhat variable among listeners and programs. Over the years I have used the prevailing favorite weighting, beginning with A, moving to B and ending up with C. As I said in my first JAES paper, Toole, 1982, even when measured levels are identical

“listeners should be interrogated on the matter of relative loudness of sounds and the levels adjusted as necessary to remove anomalies”. Forty-three years haven’t changed this.

10.5 IS IT PREFERENCE OR ACCURACY THAT IS EVALUATED?

These methods have been criticized for many reasons, and I would like to think that all have been satisfactorily put to rest. Perhaps the most persistent challenge is the belief that as long as the term “preference” is involved, the results merely reflect the likes, dislikes and tastes of an inexpert audience, not the analytical criticism of listeners who know what good sound really is. The assertion is that the objective should be “realism”, “accuracy”, “fidelity”, “truth” and the like, or even “resolving power”.

As it turns out, this is where my investigations started, 60 years ago. I required listeners to report their summary opinions, their “gestalt”, as a number on a 0 to 10 “Fidelity” scale. A 10 represented the most perfect sound they can recall, and a 0 was unrecognizable rubbish. When these tests began it is fair to say that some of the loudspeakers approached “rubbish” and none came close to perfection. Consequently, scores extended over a significant range. To help listeners retain an impression of what constituted good and bad sound, the Fidelity scale was anchored by always including in the population of test loudspeakers at least one that was at the low end of the scale and one that was at the high end, usually the one we dubbed the “king of the hill”, the best of test population to that date. A parallel scale was provided, called “preference”, thinking that there might be a difference, but the two ratings “fidelity” and “preference” simply tracked each other.

Figure 2.7 illustrates the cumulative results from our NRCC tests in the early 1980s—a slice of the history of audio. It is obvious that there are many loudspeakers clustered near the top of the scale. As time passed more products were rated highly, and listeners complained about wasting time listening to bad sound. I stopped anchoring the scale at the low end, and an interesting thing happened: most listeners continued to use the same numerical range to describe what they heard. The scores rarely went above 8, which is a common

subjective scaling phenomenon—people like to leave some "headroom" in case something better comes along. The advantage was that it was now easier to distinguish the rankings of the remaining loudspeakers because they were more widely separated on the scale. However, it was evident that they were no longer judging on the same scale—the upper end of it had stretched. When listening to just a few of the top-rated loudspeakers, which on the original scale would have been crowded together, statistically difficult to separate, the same product ratings would cover a much larger portion of the scale and be easy to separate and rank. Listeners were still judging "fidelity", but it was rated on an elastic scale—see Figure 1.4.

Listeners cannot know what all original recordings sounded like to their creators, and during evaluations they comment extensively on audible defects, rather than virtues. The least bad loudspeaker gets the highest ratings. There is a clear "preference" for loudspeakers without audible defects. What is left when the audible defects are removed? Does it have a definitive name? "Neutral" and "transparent" are logical. It could have been labeled a "neutrality" scale, but that is somewhat esoteric for most people. I chose to call it preference, but other descriptors are possible.

> Listeners obviously have a natural "preference" for high "fidelity", "accuracy", "realism", "neutrality", "transparency", "resolution" and more. It is really one scale. It is what remains when audible problems are absent.

REFERENCES

Bech, S. (1989). "The Influence of Room Acoustics on Reproduced Sound, Part 1—Selection and Training of Subjects for Listening Tests", *Audio Eng. Soc., 87th Convention*, Preprint 2850. ⏎

Bech, S. (1992). "Selection and Training of Subjects for Listening Tests on Sound-reproducing Equipment", *J. Audio Eng. Soc.*, vol. 40, pp. 590–610. ⏎

Bech, S. (1998). "Calibration of Relative Level Differences of a Domestic Multichannel Sound Reproduction System", *J. Audio Eng. Soc.*, vol. 46, pp. 304–313. ⏎

Bech, S. and Cutmore, N. (2000). “Multichannel Level Alignment, Part V: The Effects of Reproduction Level, Reproduction Room, Step Size, and Symmetry”, *Audio Eng. Soc. 108th Convention*, Preprint 5085.

Bech, S. and Zacharov, N. (1999). “Multichannel Level Alignment, Part III: The Effects of Loudspeaker Directivity and Reproduction Bandwidth”, *Audio Eng. Soc. 106th Convention*, Preprint 4909.

Bech, S. and Zacharov, N. (2006). *Perceptual Audio Evaluation*, Wiley, West Sussex, England.

Benjamin, E. (2004). “Preferred Listening Levels and Acceptance Windows for Dialog Reproduction in the Domestic Environment”, *117th Convention, Audio Eng. Soc*., Paper 6233.

Clark, D. (1981). “High Resolution Subjective Testing Using a Double-blind Comparator”, *69th Convention, Audio Eng. Soc*., Preprint 1771.

Clark, W. (1991). “Noise Exposure from Leisure Activities: A Review”, *J. Acoust. Soc. Am*., vol. 90, pp. 175–181.

Dash, I., Mossman, M. and Cabrera, D. (2012). “The Relative Importance of Speech and Non-speech Components for Preferred Listening Levels”, *132nd Convention, Audio Eng. Soc*., Paper 8614.

Fastl, H. and Zwicker, E. (2007). *Psychoacoustics, Facts and Models*, Springer, Berlin.

Gabrielsson, A., Hagerman, B. and Bech-Kristensen, T. (1991). “Perceived Sound Quality of Reproductions with Different Sound Levels”, *Karolinska Institute*, Stockholm, Rep. TA 123.

Gabrielsson, A. and Sjögren, H. (1979). “Perceived Sound Quality of Sound Reproducing Systems”, *J. Acoust. Soc. Am*., vol. 65, pp. 1919–1933.

Gedemer, L. (2015). “Subjective Listening Tests for Preferred Room Response in Cinemas—Part 1: System and Test Descriptions”, *Audio Eng. Soc., 139th Convention*.

Gedemer, L. (2016). “Subjective Listening Tests for Preferred Room Response in Cinemas—Part 2: Preference Test Results”, *Audio Eng. Soc., 140th Convention*.

Gedemer, L. and Welti, T. (2013). “Validation of the Binaural Room Scanning Method for Cinema Audio Research”, *135th Convention, Audio Eng. Soc*., Preprint 8974.

Greiner, R. A. (1980). “Amplifier-loudspeaker Interfacing”, *J. Audio Eng. Soc*., vol. 28, pp. 310–315. ⏎

ISO 226 (2003). “Acoustics—Normal Equal-loudness Contours”, *International Organization for Standardization*, Geneva, Switzerland. ⏎

ITU-R BS.1116-3 (2015). “Methods for the Subjective Assessment of Small Impairments in Audio Systems”, *International Telecommunication Union*, Geneva.

ITU-R BS.1770-5 (2023). “Algorithms to Measure Audio Programme Loudness and True-peak Audio Level”, *International Telecommunication Union*, Geneva. ⏎

Jamieson, A. (2015). “Use of and Attitudes Towards Hearing Protection in the Sound and Music Industries: Results of a Pilot Survey”, *Audio Eng. Soc., 58th International Conference*. ⏎

Lipshitz, S. (1990). “The Great Debate—Some Reflections Ten Years Later”, *8th International Conference, Audio Eng. Soc*. ⏎

Lipshitz, S. and Vanderkooy, J. (1981). “The Great Debate: Subjective Evaluation”, *J. Audio Eng. Soc*., vol. 29, pp. 482–491. ⏎

Martinez, C. and Gilman, S. (1988). “Results of the 1986 AES Audiometric Survey”, *J. Audio Eng. Soc.*, vol. 36, no. 9, pp. 686–691. ⏎

MatthewS (2023). “Blind Listening Test 2”. www.audiosciencereview.com/forum/index.php?threads/blind-listening-test-2-neumann-kh-80-vs-jbl-305p-mkii-vs-edifier-r1280t-vs-rcf-arya-pro5.43343/. Accessed August 2024. ⏎

McMullin, E., Celestinos, A. and Devantier, A. (2015). “Environments for Evaluation: The Development of Two New Rooms for Subjective Evaluation”, *139th Convention, Audio Eng. Soc*., Paper 9460. ⏎

Moore, B. C. J., Glasberg, B. R. and Baer, T. (1997). “A Model for the Prediction of Thresholds, Loudness, and Partial Loudness”, *J. Audio Eng. Soc*., vol. 45, pp. 224–240. ⏎

Nousaine, T. (1990). “The Great Debate: Is Anyone Winning”, *8th International Conference, Audio Eng. Soc*. ⏎

Olive, S. E. (1990). “The Preservation of Timbre; Microphones, Loudspeakers, Sound Sources and Acoustical Spaces”, *8th International Conference, Audio Eng. Soc*., Paper D1. ⏎

Olive, S. E. (1994). “A Method for Training Listeners and Selecting Program Material for Listening Tests”, *Audio Eng. Soc. 97th Convention*, Preprint 3893. ⏎

Olive, S. E. (2001). “A New Listener Training Software Application”, *110th Convention, Audio Eng. Soc*., Preprint 5384. ⏎

Olive, S. E. (2003). “Difference in Performance and Preference of Trained versus Untrained Listeners in Loudspeaker Tests: A Case Study”, *J. Audio Eng. Soc*., vol. 51, pp. 806–825. ⏎

Olive, S. E. (2009). “A New Reference Listening Room for Consumer, Professional and Automotive Audio Research”, *126th Convention, Audio Eng. Soc*., Paper 7677. ⏎

Olive, S. E., Castro, B. and Toole, F. E. (1998). “A New Laboratory for Evaluating Multichannel Audio Components and Systems”, *105th Convention, Audio Eng. Soc.*, Preprint 4842. ⏎

Olive, S. E., Schuck, P. L., Sally, S. L. and Bonneville, M. (1995). “The Variability of Loudspeaker Sound Quality Among Four Domestic Sized Rooms”, *99th Convention, Audio Eng. Soc*., Preprint 4092. ⏎

Olive, S. E. and Toole, F. E. (1989). “The Evaluation of Microphones—Part 1: Measurements”, *87th Convention, Audio Eng. Soc*., Preprint 2837. ⏎

Olive, S. E., Welti, T. and Martens, W. L. (2007). “Listener Loudspeaker Preference Ratings Obtained in situ Match Those Obtained Via a Binaural Room Scanning Measurement and Playback System”, *122nd Convention, Audio Eng. Soc*., Preprint 7134. ⏎

Olive, S. E., Welti, T. and McMullin, E. (2013). “Listener Preference for In-Room Loudspeaker and Headphone Target Responses”, *135th Convention, Audio Eng. Soc*., Paper 8994. ⏎

Olive, S. E., Welti, T. and McMullin, E. (2014). “The Influence of Listeners’ Experience, Age and Culture on Headphone Sound Quality Preferences”, *137th Convention, Audio Eng. Soc*., Paper 9177. ⏎

Rumsey, F., Zielinski, S., Kassier, R. and Bech, S. (2005). “Relationships between Experienced Listener Ratings of Multichannel Audio Quality and Naïve Listener Preferences”, *J. Acoust. Soc. Am*., vol. 117, pp. 3832–3840. ⏎

Self, D. R. G. (1988). “Science v. Subjectivism in Audio Engineering”, *Electronics and Wireless World*, vol. 94, pp. 692–696. www.dself.dsl.pipex.com/ampins/pseudo/subjectv.htm ⏎

Soulodre, G. A. (2004). “Evaluation of Objective Loudness Meters”, *116th Convention, Audio Eng. Soc*., Preprint 6161. ⏎

Soulodre, G. A. and Norcross, S. G. (2003). “Objective Measures of Loudness”, *115th Convention, Audio Eng. Soc*., Preprint 5896. ⏎

Suokuisma, P., Zacharov, N. and Bech, S. (1998). “Multichannel Level Alignment, Part 1: Signals and Methods”, *Audio Eng. Soc. 105th Convention*, Preprint 4815. ⏎

Toole, F. E. (1975). “Damping, Damping Factor, and Damn Nonsense”, *Audio Scene Canada*, February, pp. 16–17. ⏎

Toole, F. E. (1982). “Listening Tests—Turning Opinion into Fact”, *J. Audio Eng. Soc*., vol. 30, pp. 431–445. ⏎

Toole, F. E. (1985). “Subjective Measurements of Loudspeaker Sound Quality and Listener Preferences”, *J. Audio Eng. Soc*., vol. 33. pp. 2–31. ⏎

Toole, F. E. (1986). “Loudspeaker Measurements and Their Relationship to Listener Preferences”, *J. Audio Eng. Soc*., vol. 34, pt. 1, pp. 227–235, pt. 2, pp. 323–348. ⏎

Toole, F. E. (1990). “Loudspeakers and Rooms for Stereophonic Sound Reproduction”, *AES 8th International Conference “The Sound of Audio”*, Paper 8-011. ⏎

Toole, F. E. (1991). “Binaural Record/Reproduction Systems and Their Use in Psychoacoustic Investigations”, *91st Convention, Audio Eng. Soc*., Preprint 3179. ⏎

Toole, F. E. and Olive, S. E. (1988). “The Modification of Timbre by Resonances: Perception and Measurement”, *J. Audio Eng. Soc*., 36, pp. 122–142. ⏎

Toole, F. E. and Olive, S. E. (1994). “Hearing Is Believing vs. Believing Is Hearing: Blind vs. Sighted Listening Tests and Other Interesting Things”, *97th Convention, Audio Eng. Soc*., Preprint 3894. ⏎

Toole, F. E. and Olive, S. E. (2001). “Subjective Evaluation”, Chapter 13 in Borwick, J. editor, *Loudspeaker and Headphone Handbook*, Focal Press, Oxford. ⏎

Welti, T. and Olive, S. (2013). “Validation of the Binaural Room Scanning Method Using Subjective Ratings of Spatial Attributes”, *48th International Conference, Audio Eng. Soc*., Paper 1–2. ⏎

Zacharov, N. (1998). “An Overview of Multichannel Level Alignment”, *Audio Eng. Soc. 15th International Conference*, Paper 15–016. ⏎

Zacharov, N. and Bech, S. (2000). “Multichannel Level Alignment, Part IV: The Correlation between Physical Measures and Subjective Level Calibration”, *Audio Eng. Soc. 109th Convention*, Preprint 5241. ⏎

Zacharov, N., Bech, S. and Suokuisma, P. (1998). “Multichannel Level Alignment, Part II: The Influence of Signals and Loudspeaker Placement”, *Audio Eng. Soc. 105th Convention*, preprint 4816. ⏎

CHAPTE
R 11

Closing the Loop: Predicting Listener Preferences From Measurements

Floyd Toole

DOI: 10.4324/9781003477495-12

Psychoacoustics research scientists find models of perceptual processes very appealing. The ideal result is to be able to insert measured numbers into an equation that models a perceptual process, and accurately predict a subjective response. It has not been straightforward for individual perceptual dimensions, much less for more complex ones. Something as multidimensional and abstract as preference in the sound quality from loudspeakers would seem to present a new level of challenge—one that many people have said is impossible.

There have been a few attempts, all of which have shown some degree of correlation with perceptions. Olive (2004a) discusses a number of them. The oldest investigations suffer from a lack of truly good loudspeakers, and a lack of comprehensive anechoic measurements on them. Differences were

audible and measurable, but the differences were so large and the misbehavior so severe that almost any measurement would have shown a correlation with perceptions.

As my own research was nicely getting underway at the NRCC, *Consumer Reports* magazine began to rate hi-fi loudspeakers. It is the publication of Consumers Union (CU) that tests and rates the performance of many consumer products. The process was based on a sound power estimate compiled from anechoic measurements in 1/3-octave bands. Avoiding the most problematic low-frequency room mode region, sound levels at frequencies from 110 Hz to 14 kHz were converted to loudness units, sones, in an attempt to get closer to what might be perceived. The sone-converted curve was then compared to an "ideal" target and an accuracy score calculated. Listening tests were conducted, and results are claimed to have supported the predicted ratings, although no details of those tests appear to have been published. It seemed to be a serious-minded attempt at correlating measurements with subjective perceptions. However, the method had several problematic aspects that would gradually be revealed, resulting in revisions to the process. No scientific papers were published, only magazine articles (Seligson, 1969; *Consumer Reports*, 1973). I visited CU to see the process firsthand and was not impressed with the rigor with which the subjective opinions that supposedly justified their method had been collected. The data gathering lacked resolution, being 1/3-octave band filtered, and processing had some scientific components, but the manner in which they were employed seemed to lack justification.

As I report in detail, with references, in Section 1.1 of Part One of Toole (1986), there were several different approaches to measuring loudspeakers at that time, ranging from focusing on the on-axis response, through combinations of on- and off-axis data, total radiated sound power, and in-room listening position steady-state measurements. All had serious-minded supporters, but all could not be right. The sound power approach was supported by several respected people and companies in the eastern United States, where CU is located.

Allison (1982) criticized the method for not adequately weighting low-frequency performance, including his special interest, adjacent boundary interactions. However, on the matter of the basic method he said:

> The most important aspect of a loudspeaker's performance (certainly its most clearly audible aspect) is how its acoustic-power output varies as a function of frequency. There is wide, but certainly not universal, agreement with CU on this point. I am among those who agree.

I was interested in this process, but by 1980 it was evident from my own work that 1/3-octave measurements lack the resolution to reveal all audible subtleties, and that a single curve was not adequate to describe how a loudspeaker might sound. Also, based on direct product comparisons, there were strong disagreements between their accuracy scores and the subjective ratings that were coming out of my well-controlled listening tests, described in Toole (1982).

By the early 1980s I had created a digitally controlled, 360° data acquisition system with data post processing, calibrated the anechoic chamber to low frequencies, and had accumulated enough subjective and objective data to publish two major papers in *JAES* (Toole, 1985, 1986). Figure 2.4 shows results from the author's early investigations. They show anechoic loudspeaker measurements organized according to subjective sound quality ratings. There is no doubt that there is a relationship between the visible characteristics of the measured curves and the subjective "fidelity" ratings. The trends in the families of curves are visually evident: smoother, flatter, wider-bandwidth on-axis curves are associated with higher subjective ratings. Good performance at angles well off axis appear to be a positive contributing factor. Consistent features in the family of curves were indicators of resonances, low bass extension was an observable advantage, and so on. It was a clear indication of where future research needed to go. Visual inspection of these data allowed the recognition of potentially good or less good sounding loudspeakers from a set of anechoic measurements. It was a good start. It also showed that sound power, what CU based its accuracy score on, was, by itself, not a good predictor of

sound quality. However, when viewed as a member of the spinorama family of curves, irregularities in all of them indicated the presence of resonances, so the smoothness of the curve was relevant.

In 1985 I hosted a multiday loudspeaker measurement seminar at the NRCC, where we showed and demonstrated the measurement capabilities and had lectures on variables that were believed to be measurable and important to what we heard, and preferred. The ultimate test was a double-blind subjective evaluation—ears and brains determined what was good, not the appearance of a curve. We also demonstrated those. The audience was a who's who of loudspeaker designers and audio researchers from the United States and Canada. Among them were two engineers from *Consumer Reports*. Knowing they were coming, I had an overhead transparency prepared. For four loudspeakers that both of us had recently evaluated it showed the subjective "fidelity" rating from our double-blind listening tests, and the "accuracy score" from their publication. The correlation was −0.7 (negative 0.7!). I still have the transparency. There was discussion, some puzzlement on their part, but nothing changed at the magazine. Based on this comparison, readers would have been well advised to invert the page of ratings that they published to see something closer to the truth.

Over the years the CU loudspeaker ratings were frequently at odds with our findings, and discussions with several loudspeaker manufacturers found confirming evidence, all revealing discontent with the situation. "High-end" consumers may put more weight on recommendations from audiophile publications, but masses of general consumers paid attention to *Consumer Reports* ratings. For me it came to a head after I joined Harman International as corporate vice-president of acoustical engineering. In 1999, when some of our products received poor CU ratings I was challenged by my management colleagues to explain why. Fortunately, I knew why, and the problem was theirs, not ours.

At the time Harman was in the process of upgrading its measurement facilities and defining a process by which our many products could be evaluated. There was a confidence that our products were good performers,

but there were exceptions. This situation provided an opportunity to move towards a trustworthy performance evaluation process within Harman, and possibly beyond. My suggestion was to initiate research aimed at finding a better way of rating products that we all could use, which, if successful, might be enough to persuade CU to change their evaluation method. It would require time, manpower and money. Remarkably, the CEO, Bernie Girod, and Dr. Sidney Harman, corporation chairman, agreed. This was an important decision, and these men deserve credit for having faith in the scientific method.

My long-term colleague from NRCC days, Sean Olive, was then in my research group at Harman. We assessed where we stood with respect to psychoacoustic rules that could be applied to numerically evaluating spinorama curves to reveal subjective preferences. A lot was known about the visible clues in the comprehensive anechoic data, including the following desirable properties:

- Smooth and flat on-axis and listening window (direct sound) responses
- Similarly smooth (but not necessarily flat) listening window, early reflections and sound power curves, indicating a lack of annoying resonances—the dominant degradation of neutral timbre
- Bass extension.

It seemed that a computer model could be provided with basically useful data, and then one would see what insights it could add. Sean then initiated the following exercise, all of which is described in Olive (2004a, 2004b). In parallel with this, the spinorama measurement technique was being finalized, and it would play a significant role.

11.1 THE OLIVE EXPERIMENTS—PART ONE

Using the spinorama collection of anechoic measurements, an evaluation of 13 loudspeakers that had recently been reviewed by CU was undertaken. All were bookshelf-sized products, and all lacked low bass. It began with a fully balanced, double-blind, listening test (every loudspeaker was auditioned against all others the same number of times) by a group of selected and trained listeners. The result is shown in Figure 11.1(a) compared with the corresponding accuracy scores for the same loudspeakers published by CU. The contrast between the two sets of data is striking, with loudspeaker 1 exhibiting both "best" and "worst" evaluations. Overall, there was no correlation between the two results, indicating that, if the subjective data are correct, the CU method of evaluation is based on a faulty premise: that a sound power measurement alone contains information necessary to describe listener perceptions of sound quality in small rooms.

Employing the much more comprehensive, higher-resolution data generated routinely in the Harman International tests, Olive identified a model that described the subjective preference ratings in a near-perfect manner, as shown in Figure 11.1(b). The accuracy exhibited by this model implies a complex analysis, and this is indeed the case.

The statistics derived from the different curves of the anechoic spinorama data are as follows:

- AAD: *absolute average deviation* (dB) relative to the mean sound level between 200 and 400 Hz, in 1/20-octave bands from 100 Hz to 16 kHz
- NBD: average *narrow-band deviation* (dB) in each 1/2-octave band from 100 Hz to 12 kHz
- SM: *smoothness* (r^2) in amplitude response based on a linear regression line through 100 Hz to 16 kHz
- SL: *slope* of the best-fit linear regression line (dB)

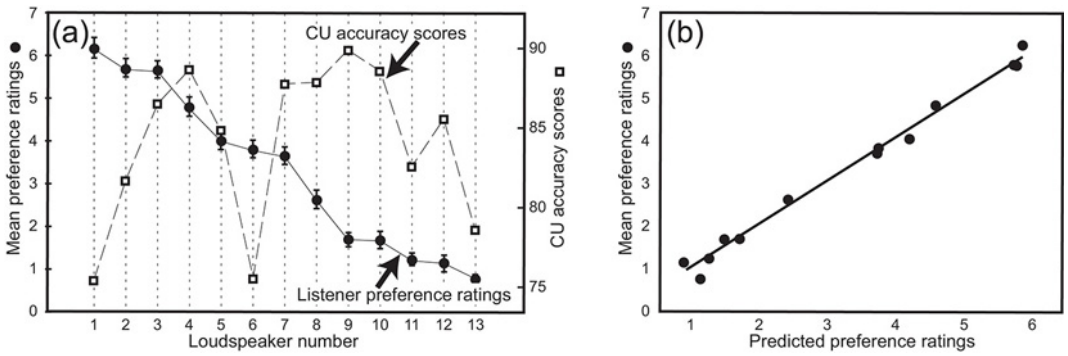


FIGURE 11.1 *(a) shows a comparison of mean preference ratings from double-blind listening tests and accuracy scores from evaluations of the same loudspeaker models published by Consumers Union (CU). The correlation between the two was small and negative ($r = -0.22$) and not statistically significant ($p = 0.46$). Confidence intervals of 95% are shown for the subjective preference ratings, indicating high confidence in several of the rating differences, although there were clusters of loudspeakers that were in statistical ties: 2 and 3; 5, 6 and 7; 9 and 10; and 11 and 12. Data from Olive (2004a, figure 3). (b) shows the comparison of the same subjective data and predictions from a model created by Olive, using anechoic data of the kind shown in Figure 2.9. In this case the correlation was 1.0 ($r = 0.995$) and the statistical significance a very high $p > 0.0001$. Data from Olive (2004b, figure 4). It is interesting that loudspeaker number 1 was a Harman product—best in our test, worst in theirs.* ⏎

These statistics were applied to all the measured spinorama curves (Figure 2.9):

- On-axis (ON)
- Listening window (LW)
- Early reflections (ER)
- Sound power (SP)

- Both directivity index curves (ERDI and SPDI)
- Predicted-in-room (PIR), a proportioned combination of ON, ER and SP that approximates measured room curves in several typical rooms

Then there are two statistics that separately focus on low-frequency performance:

- LFX: low-frequency extension (Hz), the 6 dB down point relative to listening window (LW) sensitivity in the range 300 Hz to 10 kHz
- LFQ: absolute average deviation (dB) in bass frequency response from LFX to 300 Hz

When the dominant contributing factors to this model were analyzed, they were:

- Smoothness and flatness of the on-axis frequency response: 45%
    - AAD_ON (18.64%) + SM_ON (26.34%)
- Smoothness of sound power, SM_SP: 30%
- Bass performance: 25
    - LFX (6.27%) + LFQ (18.64%)

On-axis flatness, and on- and off-axis (sound power) smoothness account for 75% of the preference estimate. These are indications of a lack of resonances. Bass performances alone accounts for 25% of the preference estimate, something not to be dismissed. Putting it in different terms:

1. Smoothness and flatness in *both* on-axis and sound power indicate an absence of resonances—timbral neutrality.
2. Smooth sound power combined with smooth and flat on-axis response is an indicator of smoothly changing directivity index as a function of

frequency. Less timbral degradation of reflected sounds. Numbers 1 and 2, representing timbre, combine to represent 75% of the total rating.

3. Bass extension and smoothness are fundamentally important qualities, representing 25% of the total rating. The limited bandwidth of these bookshelf loudspeakers would not strongly energize very low-frequency room resonances, so we are seeing evidence of loudspeaker performance.

None of this could be seen in the measurements used by CU.

These data were shared with CU before public presentation, and they acknowledged that they had problems, and agreed to further discussions. Most importantly, they ceased evaluating hi-fi loudspeakers. Harman International did a favor to consumers and the loudspeaker industry as a whole. There were a few discussions, and it seemed that improvements were being made to their process, but eventually the personnel involved moved on to other projects.

11.2 THE OLIVE EXPERIMENTS—PART TWO

As impressive as this result is, there is a common factor that prevents the model from being generalized—all 13 of the loudspeakers were bookshelf models, a natural basis for a comparative product review in a magazine. All suffered from a lack of low bass, and many of them had common features: enclosure size, driver complement and configuration, etc. The population needed to be expanded to include larger products, using different driver configurations and types. This was done in a second test that involved 70 loudspeakers from many origins, covering large price and size ranges. The penalty in this test was that the listening had been done in a fragmented manner, 19 independent tests over an 18-month period, as happens in the normal course of business. All loudspeakers were evaluated in comparison with some other models, but there was no overall organization to ensure that the comparisons were balanced, each model against all other models. It would be unreasonable to expect the same remarkable accuracy in the subjective-objective correlations as has just been seen.

Still, the result was impressive: predicted preference ratings correlated with those from listening tests with a correlation of 0.86, with a very high statistical significance ($p > 0.0001$). These are outstanding numbers given the opportunities for variation in the listening tests, meaning that the listeners themselves were highly stable "measuring instruments" and that the strategy of always doing multiple (usually four products) comparisons is a good one. The model that, for these products and these subjective data produced the best result was different from the earlier one.

Here, the weighting given to the factors in predicting preference were:

- Narrow-band and overall smoothness in the on and off-axis response: 38%
    - NBD_PIR (20.5%) + SM_PIR (17.5%)
    - Since PIR incorporates on-axis, early-reflection and sound power curves it is evident that all three must exhibit narrow and broadband smoothness.

- Narrow-band smoothness in on-axis frequency response (NBD_ON): 31.5%
- Low-bass extension (LFX): 30.5%

Again, to explain in plain language: PIR is an empirical prediction of a steady-state room curve based on 12% listening window, 44% early reflections and 44% sound power (Devantier, 2002). In this case, my opinion is that in-room response, by itself, has not taken on a special importance. Instead, the model has recognized NBD and SM, both indicators of smoothness in the data from which PIR is calculated. When such smoothness is seen in all three of these curves it is a strong indicator of an absence of resonances. Smoothness is also seen in the on-axis (direct sound) curve. Combined, indicators of timbral neutrality accounted for 69.5% of the total score. Bass extension is highly valued, at 31.5% of the total score. Bass smoothness, LFQ, was not consequential in this model, probably because many of these loudspeakers had the bass energy to excite invasive room resonances, that would have masked any intrinsic smoothness the loudspeakers may have had.

11.3 THE OLIVE EXPERIMENTS—PART THREE

The following was an inspired addition to the Part One tests. Listeners were required to "draw" (using sliders on a computer screen) a frequency response, describing what they thought they heard in terms of spectral balance. This is a task that obviously could not be asked of average listeners, but these listeners had been through a training program (Olive, 1994, 2001) and were able to estimate the frequencies at which audible excesses and deficiencies occurred. When these data were compiled, the results indicated that all the highly rated loudspeakers had been judged to have very flat curves. Of great interest, of course, is which one of the several measured spinorama curves for these loudspeakers does this correspond to—the logical candidates being: sound power (the CU model), the early-reflections curve which is similar to the PIR (the steady-state room curve that is the basis of room-equalization schemes) or the on-axis curve (not the dominant measured component in rooms, but important in perception)?

Figure 11.2 shows a revealing comparison of three loudspeakers chosen because of their ratings in Figure 11.1. The subjectively "drawn" spectrum curves are obviously best matched to the on-axis and listening window frequency responses of the loudspeakers, thus confirming the observation noted from the earliest tests shown in Figure 2.4, that a good axial amplitude response, delivering the direct sound to listeners, appears to be a requirement for good sound. Beyond that, attention must be given to off-axis behavior.

Let us dig deeper into these examples, to see why the Harman and CU results did not always agree.

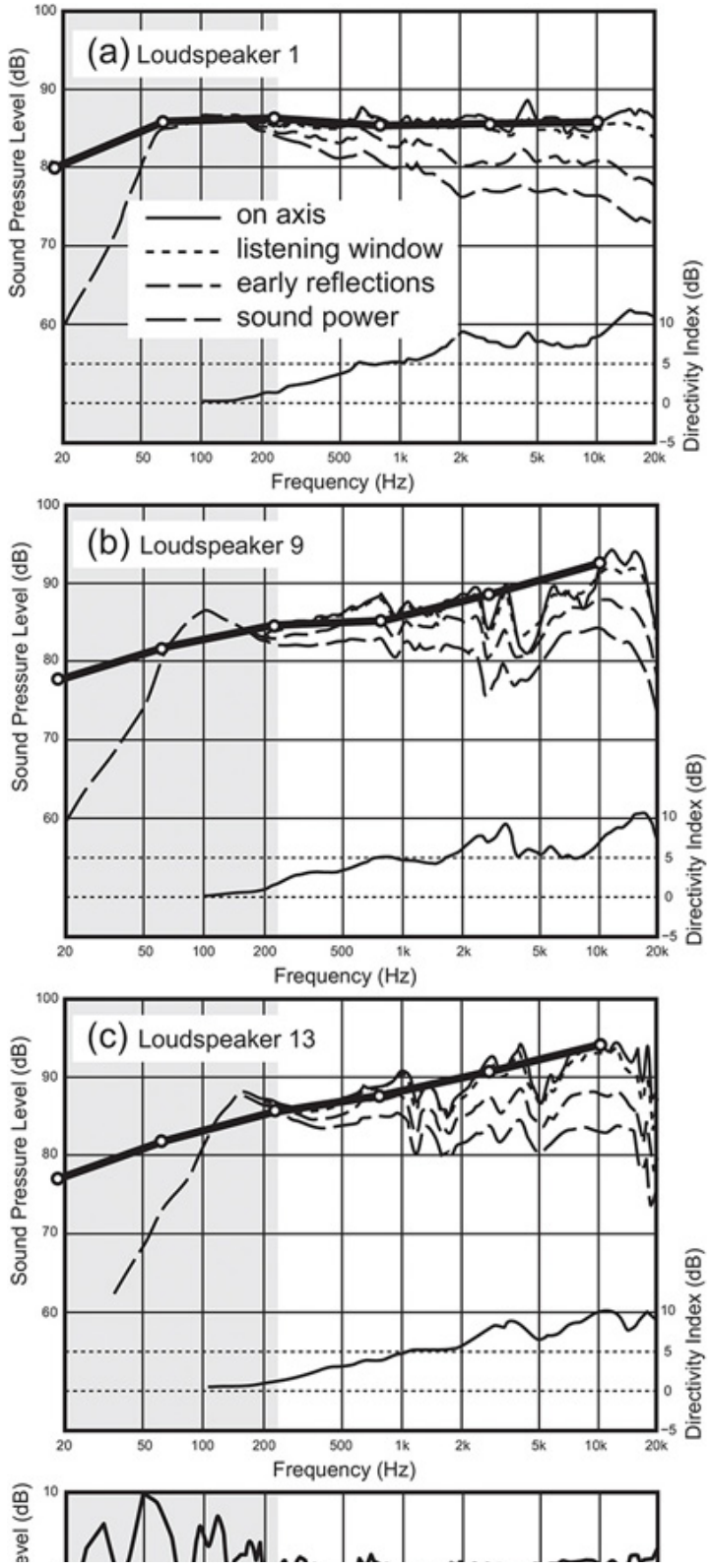
- Loudspeaker 1, the highest rated in Harman tests, and the lowest rated in CU tests. A straightforward examination of the spinorama reveals a wide bandwidth, flattish and quite smooth on-axis amplitude response. Bass extends smoothly down to about 60 Hz. The smoothness is maintained in the progressive spatial averages, indicating an absence

of coloring resonances. The DI is respectably smooth. This should be a good sounding, highly rated loudspeaker, and it is revealed as such in the Harman double-blind listening tests. However, the CU calculated "accuracy" clearly did not approve, possibly because of the tilted sound power curve.

- Loudspeaker 9, the highest rated in CU tests and one of the low rated ones in the Harman tests. The rising axial response is a bad start, and the severe resonances seen in all of the curves indicate significant coloration. The bass shows a bump around 100 Hz, lending "punch" to kick drums, but it does not extend as far down as Loudspeaker 1. DI indicates some problems in the crossover region between the woofer/midrange and tweeter. The low subjective rating in Harman evaluations seems justified. The CU calculation appeared to like the flatter sound power. The 1/3-octave resolution in the CU sound power data would not reveal the resonances, and the lack of high-resolution anechoic data hinders the identification and evaluation of resonances.
- Loudspeaker 13, the lowest rated in the Harman tests, and also low rated in CU tests. Again a dramatically rising on-axis response is not good. Evidence of several resonances is not good. A bass roll-off that begins around 120 Hz, following a bass bump, is a significant deficiency—bass extension is important. This is not a good sounding loudspeaker, and both evaluations identified it as such.
- A question to ponder: were some of these bad sounding loudspeakers designed deliberately to score highly in CU evaluations?

At low frequencies the anechoic data cease to reflect accurately what is heard in the listening room—it will be less smooth and significantly higher in amplitude in the room, and this portion of the curves is shaded. As the tests were done using positional substitution, the contribution of the room to the sound would be quite constant. (d) shows the average error between predicted in-room (PIR) response (from anechoic data) and real steady-state

in-room measurements (from Olive, 2004b, figure 11.7). The room dominates what is heard below about 250 Hz.



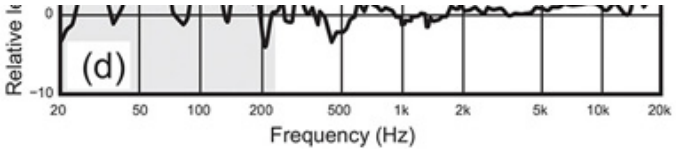


FIGURE 11.2 *Figures (a), (b) and (c) show the spinorama data for three of the 13 loudspeakers tested (results in Figure 11.1). The heavy dark line superimposed on each curve is the subjective estimate of the spectral balance given by the listeners while in the listening room. The open circles indicate the center frequencies of bands used by listeners to describe the spectra of what they heard. This is a "dimensionless" curve on the vertical axis, reflecting subjective impressions of boost and cut relative to the personal ideals of the listeners. Consequently, this author has taken the liberty of adjusting the vertical scale and the vertical position of these curves until they appeared to fit the pattern of anechoic data. The chosen vertical scale is used in all three curves. The most obvious visual correlation with the subjective spectral balance is with the anechoic on-axis and listening window curves.* ⏎

In these experiments trained listeners were able to draw curves of spectral trends—crude frequency responses—describing what they heard in the listening room. All the high-scoring loudspeakers were described as having flattish spectra (above the transition frequency), a trend that matches the flattish on-axis/listening window curves for all of the corresponding anechoic measurements.

Loudspeakers with lower preference ratings were described as having either insufficient bass or excessive treble, or both. These trends tend to yield a flatter sound power curve, and thereby higher accuracy scores in CU reviews. However, it is a trend that runs contrary to the interests of high sound quality. The direct sound has a special place in the hierarchy of sounds arriving at the ears, and flat is a good objective. Figure 2.4 indicates this, and Queen (1973) concludes an investigation by saying: "The results thus far tend to confirm the presupposition that spectrum identification

depends principally upon the direct sound of the sound source." These remarkable results further confirm it.

Below the transition frequency, in the gray shaded area in Figure 11.2, it is evident that listeners responded to something closer to the low-frequency response in the room than to the anechoic frequency response. This is the ultimate limitation of all such models, in that the listening room and the arrangements within it determine the bass level and quality. With bass accounting for about 30% of the overall subjective rating, this is a matter of great importance and it implies that if there is to be parity among listening situations—professional and consumer—there needs to be substantial control over the low frequencies in the room.

11.4 SUMMARY

Earlier in the book it was shown that anechoic frequency-response data can be used to predict steady-state room curves above the transition frequency. We know that measuring a room curve with a microphone is not the same as perceptions by the human binaural hearing system, but the information used by human listeners is hidden within the room curve. Separating the individual direct and early reflected sounds in a steady-state room curve after the fact is not possible. However, the reverse is possible. The anechoic loudspeaker measurements and the post processing for spinorama presentation was an attempt to evaluate the direct and reflected sounds arriving at a listener's ears in a typical domestic room. What Parts One and Two of this modeling exercise proved is that it succeeded sufficiently well to enable useful predictions of preference ratings of sound quality. It is confirmed that resonances are the primary problem in loudspeakers. It is also learned that bass alone accounts for approximately 30% of overall sound quality ratings. As this is dominated by room resonances, it is clear that this is a problem that must be addressed in every individual listening room.

Part Three was an outstanding act, only possible because of well-trained listeners.

Traditional psychoacoustic experiments have examined listener responses to different sound spectra by asking questions about perceptual dimensions: brightness, darkness, shrillness, and countless other adjectives people gravitate to in describing what they hear. It is a tedious semantics exercise, and there are large variations in the results as displayed in diagrams showing statistical interactions among the perceptual dimensions and measurement axes. Broad trends can be seen but little else.

These tests were totally different. Here, Olive has shown that listeners trained in recognizing resonances, the building blocks of the sounds we are interested in hearing, can describe, in physically measurable terms, why they prefer certain sounds. They told us that what they heard was best described by the on-axis and listening window curves—the direct sound. This is the "engineering" approach, and it leads directly to the technical information that matters most. With that knowledge multiple perceptual dimensions can be liberated without coloration. Well done.

The spinorama measurement and post-processing scheme (Section 2.7) that had been developed while Sean Olive's research was in progress was validated, which added to confidence in using and interpreting it. It is now incorporated in an audio industry standard, ANSI/CTA-2034-A (2015), and is used worldwide.

Conclusion: People like timbrally neutral loudspeakers with deep bass; both models say so. We thought we already knew it, but it was reassuring to have it statistically confirmed. In doing so, an influential publisher of unreliable loudspeaker reviews was silenced. Altogether a worthy effort. We may never eliminate the need for subjective evaluations, but now there are validated measurements to guide the design of loudspeakers that please listeners, and to provide insight during their evaluation.

REFERENCES

Allison, R. (1982). "Rating Loudspeakers for "Accuracy", *Stereo Rev.*, vol. 47, no. 6, pp. 60–61. ⏎

ANSI/CTA-2034-A (2015). “Standard Method of Measurement for In-Home Loudspeakers”, *Consumer Technology Association, Technology and Standards Dept*. www.CTA.tech. ⏎

Consumer Reports (1973). “Loudspeaker Accuracy: CU’s Tests”, *Consumer Reports*, vol. 38, pp. 456–457, July. ⏎

Devantier, A. (2002). “Characterizing the Amplitude Response of Loudspeaker Systems”, *113th Convention, Audio Eng. Soc*., Preprint 5638. ⏎

Olive, S. E. (1994). “A Method for Training Listeners and Selecting Program Material for Listening Tests”, *Audio Eng. Soc. 97th Convention*, Preprint 3893. ⏎

Olive, S. E. (2001). “A New Listener Training Software Application”, *110th Convention, Audio Eng. Soc*., Preprint 5384. ⏎

Olive, S. E. (2004a). “A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 1—Listening Test Results”, *116th Convention, Audio Eng. Soc*., Preprint 6113. ⏎

Olive, S. E. (2004b). “A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 2—Development of the Model”, *117th Convention, Audio Eng. Soc*., Preprint 6190. ⏎

Queen, D. (1973). “Relative Importance of the Direct and Reverberant Fields to Spectrum Perception”, *J. Audio Eng. Soc*., vol.21, pp. 119–121. ⏎

Seligson, L. (1969). “How Consumers Union Tests Loudspeakers”, *Stereo Review*, February, pp. 62–66. ⏎

Toole, F. E. (1982). “Listening Tests—Turning Opinion into Fact”, *J. Audio Eng. Soc*., vol. 30, pp. 431–445. ⏎

Toole, F. E. (1985). “Subjective Measurements of Loudspeaker Sound Quality and Listener Preferences”, *J. Audio Eng. Soc.*, vol. 33. pp. 2–31. ⏎

Toole, F. E. (1986). “Loudspeaker Measurements and Their Relationship to Listener Preferences”, *J. Audio Eng. Soc*., vol. 34, pt. 1, pp. 227–235, pt.

2, pp. 323–348. ⏎

CHAPTE
R 12

Room Acoustics and Acoustical Devices

Floyd Toole

DOI: 10.4324/9781003477495-13

Instinct might tell us that there is little in common among home listening rooms, home theaters, recording control rooms, movie dubbing stages and cinemas, but there is. They are all points on a continuum of physical dimensions and their perceptual consequences (Figure 3.3). Some of the measures we use to describe circumstances and events in these spaces have their origins in concert hall acoustical investigations, a long way from the present context. But when we look closely at the sounds heard by listeners in most sound reproduction venues, there are clear parallels. There is much to discuss.

12.1 REVERBERATION TIME

In large performance venues, concert halls and auditoriums, reverberation is real and necessary. It consists of sounds from voices and musical instruments, with limited sound power output, that are reflected many times from many surfaces, penetrating to all parts of the room, gradually decaying in amplitude with time until they are inaudible. This is what Sabine experienced, and measured by ear, when he initiated the science of room acoustics in the 1890s.

Reverberation time (RT or RT_{60}) is measured as the time for a 60 dB reduction in amplitude. There are other versions of RT measurements aimed at clarifying specific performance aspects. Performance halls need reflections to keep loudness high throughout a large audience, but enough acoustical damping must be present to prevent the reverberation from interfering with the intelligibility of voices or clarity of the music. It is a difficult compromise. Typical values for RT in performance venues are in the range 1 to 2 seconds, opera houses to concert halls, optimized for the music most commonly performed. Some halls have physically or electronically adjustable acoustics.

In contrast, typical reverberation times for domestic listening rooms and recording control rooms are typically in the range 0.2 to 0.4 seconds. These spaces have a lot of absorption and large scattering objects (furniture and equipment) relative to their size, so a classic, relatively diffuse, reverberant sound field does not exist. It is reduced to a relatively small number of reflections that progressively decay in amplitude. The CEDIA/CTA—RP22 (2023) recommendation takes the step of calling it RdT, Reflection Decay Time. It is logical, but unnecessarily confusing because measuring instruments and algorithms will persist in calling it RT or RT_{60}. This book will call it RT for simplicity, even though we know what it measures. But, if one is rigorous about names, RT measurements at low frequencies are probably measuring the ringing decay time of strong room resonances. A third name is needed. In fact, getting a trustworthy measure of RT at low frequencies can be difficult because of background noises, beating room modes, and positional variations of standing waves. There is good reason why, in Appendix D of the CEDIA/CTA document, describing the

calculation of RT_{60}, the calculations stop at the 125 Hz octave band. Below the transition frequency discrete reflections progressively cease to be a useful concept.

The high energy reverberation in classical music, many popular recordings, and movies overrides that of the listening room and, in general, listeners tend to perceive that dominance, especially in multichannel sound reproduction (Hughes et al., 2016).

Figure 12.1(a) shows the progression in RT reduction as room treatments and furnishings are added in a typical domestic room, as exemplified by the original NRCC listening room. It is seen that adding an acoustically absorbent carpet has a major effect. Adding multidirectional reflections, redirection, and scattering from cabinets causes that absorbent to work harder. Finally, adding the absorption and dispersion of normal furnishings brings the RT into the desirable range. Low-frequency membrane absorbers were incorporated because the original room was a laboratory space with concrete floor and plaster on block walls (to contain any mad-scientist experiments that might go wrong). The resulting RT is consistent with data on typical domestic rooms. From this it is clear that, with a bit of luck, a well-furnished domestic room can be a good basis for a listening environment, without additional acoustical devices.

Figure 12.1(b) shows RT data on a selection of rooms, most of which were constructed for listening evaluations on loudspeakers and other audio products. Most fall into the mid-frequency RT range 0.3 to 0.5s. Data for those rooms mingle with average data for Canadian and UK living spaces, which suggests that the results of evaluations done in those environments may translate well into the world of consumers.

Figure 12.1(c) shows target ranges for RT from some international recommendations. The IEC 60268-13 (1998) document is intended to provide guidance for conducting listening tests on loudspeakers, and presumably for enjoying them. ITU-R BS.1116–3 is aimed at a very different audience. The introduction states: “This Recommendation is intended for use in the assessment of systems which introduce impairments so small as to be undetectable without rigorous control of the experimental

conditions and appropriate statistical analysis." In other words, this is *not* what is needed for evaluating loudspeakers, or for being entertained by them. Nevertheless, the document often appears as one of the performance requirements for rooms used for general audio evaluations. EBU Tech 3276–2nd edition (1998) and EBU Tech 3276-E (2004) address the very different needs of broadcasters needing standardization of their listening circumstances. The loudspeaker and listening room requirements are identical to the ITU document.

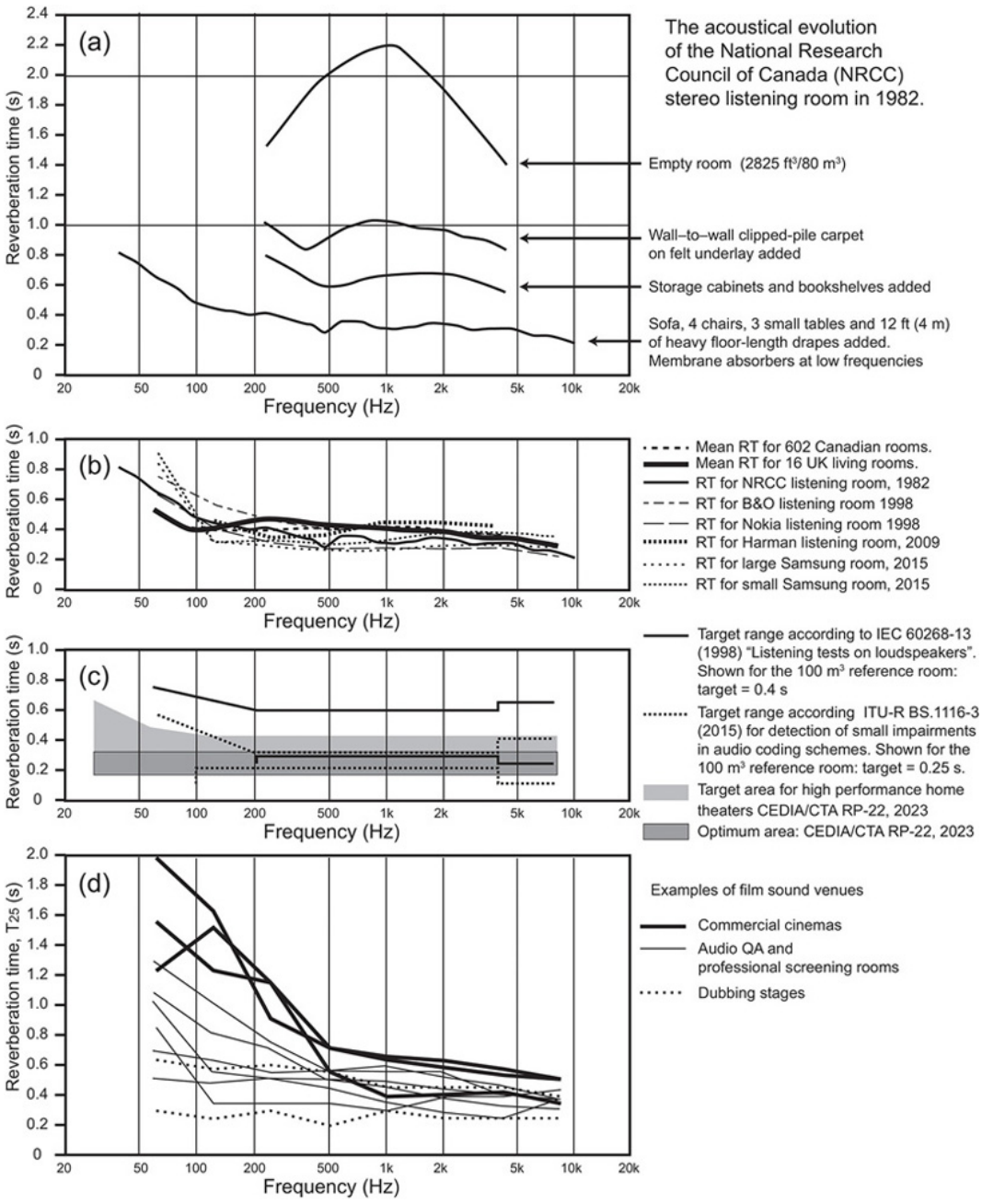


FIGURE 12.1 *(a) An illustration of how common home furnishing items can result in a good listening space. (b) RT averages for surveys in Canadian and UK homes, and example performances for several professionally constructed listening rooms. (c) The recommended ranges for RT from the noted IEC, ITU and CEDIA/CTA documents. Reproduced with permission. (d) RT data for typical cinemas, a reference cinema and two dubbing stages. From Toole (2015, figure 3) and SMPTE TC-*

25CSS (2014) reproduced with permission. Screening room data courtesy of Linda Gedemer. ⏎

Dolby makes no requirements for facilities for Atmos mixing, although they mention BS.1116 for "guidance and insight". The CEDIA/CTA recommendation followed the lead of BS.1116 for custom home theaters. These systems are designed to sound good, distinctive, and to impress, so low RT recommendations are not surprising. That said, the "optimum area" RT recommendation is about 0.2 s to 0.3 s maintained down to the 31.5 Hz octave band, the lowest frequency of which is 22 Hz. This requires expensive low-frequency absorbers, but the real problem, resonances, has other solutions (see Chapter 14).

Figure 12.1(d) shows measurements from several venues where movie sound tracks are created, inspected for excellence or experienced by customers. It is not surprising that budgetary constraints compromise the acoustical performance of mass-market cinemas, with low-frequency reverberation times clearly elevated. However, several of the audio quality assessment and screening rooms and the dubbing stages where the sound tracks are mixed fall in the range of the quality listening rooms in (b) and would be considered acceptable by IEC recommended RT values in (c). If one is looking for a concurrence of objectives for movie sound and home theater sound this would seem to be a good beginning. Just note that the auditory experiences in the local cinema multiplex may be less impressive than in good home theaters or in the dubbing stages where the films were mixed.

Cinemas exhibit mid-frequency RTs of 0.3 s to 0.6 s for good dialog intelligibility, and so that the acoustics of the often-large venue displayed on the screen will can be convincingly delivered. In terms of speech intelligibility, anything below 0.5 s is considered to be safe, but the elevated RTs seen at low frequencies can amplify lower vocal formants, causing chestiness.

Rooms as acoustically "dead" as these don't have a traditional diffuse reverberant sound field. It cannot exist. What is measured as RT is a

decaying sequence of relatively discrete reflections, not a dense, diffuse sound field.

12.1.1 Measuring Reverberation Time

Professional acousticians measuring the RT of performance venues follow highly ritualized, internationally standardized, procedures. In general, they involve a sound source that radiates omnidirectionally—a special loudspeaker, a starter's pistol, or a popping balloon—that is located where the orchestra or band plays. It is assumed that an orchestra is an approximately omnidirectional radiator. Measurements are made at several locations in the audience area using an omnidirectional microphone, and the results considered in the context of how the venue reverberation is likely to affect musical performances or speech.

For sound reproduction the process needs to be modified because the sound sources are the reproduction loudspeakers, and the venue reverberation is not expected to be a significant factor in the delivery of the program, whether it is music or movies. In fact, the principal consideration is well summed up in the requirements for good dialog intelligibility. In a home theater or cinema situation, the logical loudspeaker to provide the stimulus is the center channel, where virtually all the dialog originates, and the desirable RT is less than about 0.5 s at middle and high frequencies. This is an easily achievable objective, as is seen in Figure 12.1.

For practical reasons measurements are usually made from 5 dB after cessation of sound to 35 dB down, thus avoiding background noise and the need to produce very loud starting sounds. This is called T30, but a linear decay is assumed and it is doubled to get the time to decay 60 dB. A T20 version also exists, but they are all presented as 60 dB decay times. Therefore: T60 = T30 = T20, just measured in different ways. Obviously, if the decay is not linear, the estimate is wrong.

12.1.2 Calculating Reverberation Time

In large highly reflective rooms the reverberation time is often well predicted by the original Sabine formula:

$$RT = .049\ V/A$$

where V is the total volume in ft^3 and A is the total absorption in the room in sabins. The total absorption, A, is calculated by adding up all of the piecemeal areas (carpet, drapes, walls, etc.) of the boundaries multiplied by their individual absorption coefficients:

$$A = (S_1\alpha_1 + S_2\alpha_2 + S_3\alpha_3 \ldots)$$

Where S is the area in square feet and α is the absorption coefficient for the material covering that area. Absorption coefficient is ideally a measure of the proportion of randomly incident sound power that is absorbed by a material. The product of S and α is a number with the unit sabins. The absorption of some items, such as people or chairs, is sometimes quoted directly in sabins.

The metric equivalent of the Sabine formula is RT = 0.161V/A, where the volume is in m^3, areas are in m^2 and A is in metric sabins.

As rooms get more absorptive, smaller, and as the materials on the room boundaries become less randomly distributed (e.g. wall-to-wall carpet on the floor) this equation becomes progressively less reliable. Over the past 100 years several increasingly more complex equations have been developed in order to accommodate asymmetry in rooms and the fact that the sound field is not diffuse, among them Fitzroy (1959) and Arau-Puchades (1988). However all of them, in order to be practical, make assumptions. Dalenbäck (2000) says, "these two formulas give a better estimate than the classical formulas [Sabine and Eyring] in *some cases* but here a central question is: *how can one be sure they are better in a particular case?* So far no equation with universal applicability has been shown" (his emphasis). Fortunately, as will be seen, in small rooms for sound reproduction, high accuracy is not required. The simple Sabine formula provides estimates that are adequate for our purposes.

12.1.3 Is There a More Useful Metric for Our Purposes?

As useful as RT is for evaluating highly reflective venues, there may be a better metric for our purposes: cumulative energy time—the time taken after the arrival of the direct sound for the sound field to rise to the steady-state level. This is a much shorter time than the corresponding RT, and seems to more directly address the perceptual processes when brief sounds are involved. For example, Figure 12.2 shows that in a cinema with a 2.5 s (2500 ms) RT at 50 Hz, a level within 2 dB of steady state is reached in 90 ms. At 500 Hz, RT is 800 ms, and the cumulative energy time is 25 ms. These are enormous differences, and although RT is a related parameter, it does not embrace all temporal events that are likely to matter. Events very early in the RT decay data could be relevant, but a new form of interpretation would be required.

Germane to this discussion are the findings that for speech intelligibility —a crucial consideration for movies—it is the early reflections that are the main contributors. Early reflection energy arriving within about the first 50 ms following the direct sound has the same effect on speech intelligibility scores as an equal increase in the direct sound energy. This was true for both normal and hearing-impaired listeners. Bradley et al. (2003) say: "Although it is important to avoid excessive reverberant sound, adding large amounts of absorption to achieve very short reverberation times may degrade intelligibility due to reduced early reflection levels." They suggest a ratio of the energy within the first 50 ms of an impulse response to the energy associated with the direct sound as a new figure of merit for a room, called the early reflection benefit (ERB).

Leembruggen (2015) showed several examples in which cumulative energy data appeared to relate to details in frequency-domain measurements in ways that RT did not, and thereby, possibly to perceptions. In the context of concert halls, Bradley et al. (1997) found that the perceived strength of bass sounds was related to early and late reflection arrivals, and not significantly related to low-frequency reverberation time.

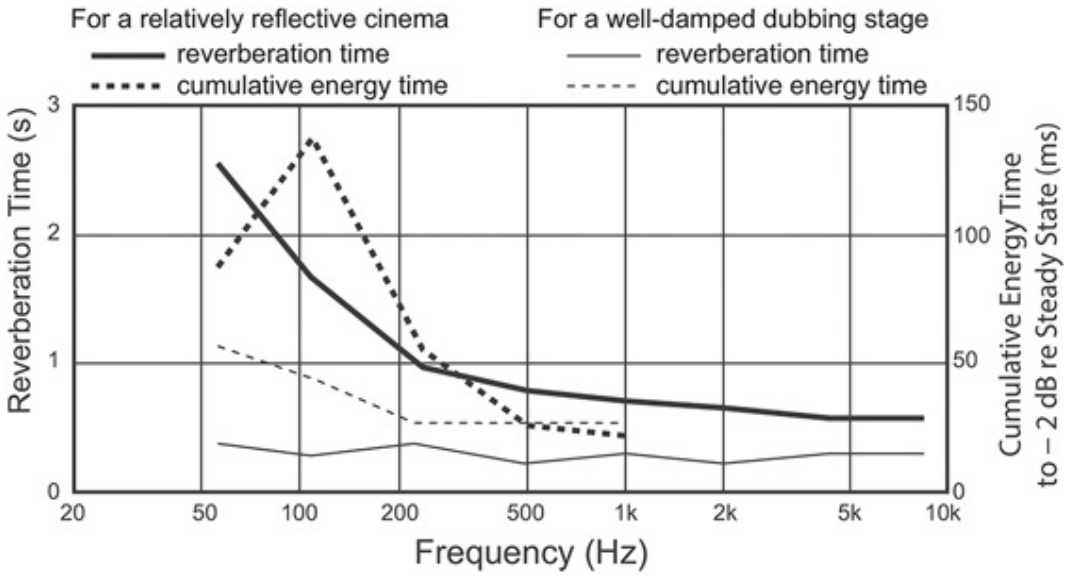


FIGURE 12.2 *A comparison of reverberation time and cumulative energy time for two very different movie playback venues. Data from Toole (2015, figures 8 and 9), based on SMPTE TC-25CSS (2014).* ⏎

Multiple factors are at play in this situation, but there are several persuasive reasons to pay special attention to early reflections, and to consider alternatives to RT as criteria of acoustical performance in sound reproduction venues.

12.1.4 Diffusion in Small Rooms

A perfectly diffuse sound field is isotropic: at any point within the sound field sounds may be expected to arrive from all directions with equal probability. It is also homogeneous: it is the same everywhere in the space. True diffusion exists only as an academic ideal. Reverberation chambers used to measure the absorption of acoustical materials are designed to be diffuse and can come close, but as soon as a test sample of absorbing material is introduced into the space, it ceases to be—contributing to variations in sound absorption coefficients measured in different labs. The assumed "random incidence" sound field is corrupted.

Diffusion can be improved by using sound scattering or dispersing objects, irregular, curved and angled surfaces, and specially designed devices, called diffusers. Perceptually, the relatively diffuse sound field in a concert hall sounds spacious and enveloping. However, a diffuse sound field is not a requirement for the perception of spaciousness and envelopment. Section 7.2 explains that much simpler sound fields created in multichannel sound reproduction can deliver sounds to the ears that are perceived to have those spatial qualities. In that situation, the listening space becomes a secondary factor.

When people talk about diffusion in small listening rooms, they misspeak. One hears of "diffusing" loudspeakers when what is meant is a widely dispersing loudspeaker that radiates its sound in many directions. We talk of "diffusers", but these devices really scatter or disperse incoming sounds in many directions and can only contribute to a diffuse sound field in a relatively reflective, live room (the room described in Section 12.5.3 is a unique exception).

If diffusion is not a requirement for listening rooms, why do we use diffusers? They are, or should be, used to attenuate specific reflections by sending most of the energy in the incident sound off in many directions. Whether the scattered sound is a problem is a reasonable concern. A diffuse sound field is not a desirable feature in a sound reproduction venue; the multichannel program should deliver the effect as needed. Fortunately, because of the amount of absorption in these spaces, it cannot exist.

Absorbing the first reflections in a room not only absorbs that sound "ray", but it prohibits any further reflections of that sound, and so the entire reflected sound field is attenuated. Any potential reverberation is severely inhibited. Listeners would be in a strong direct sound field if only first reflections were absorbed. To visualize this situation, Figure 12.3 illustrates a situation for a home listening room employing conventional forward-firing loudspeakers.

It is often instructive to put yourself at the loudspeaker location and imagine what happens to the radiated sound. Although every room will be different, it is a safe generalization that a high percentage of the mid- to

high-frequency energy is absorbed at the first surface it encounters. This is also true in cinemas, where the loudspeakers are significantly directional, being designed to deliver sound to the (highly absorptive) audience area and not to the ceiling and side walls. This helps to explain why the mid- to high-frequency RT of cinemas is comparable with that of domestic venues, as seen in Figure 12.1.

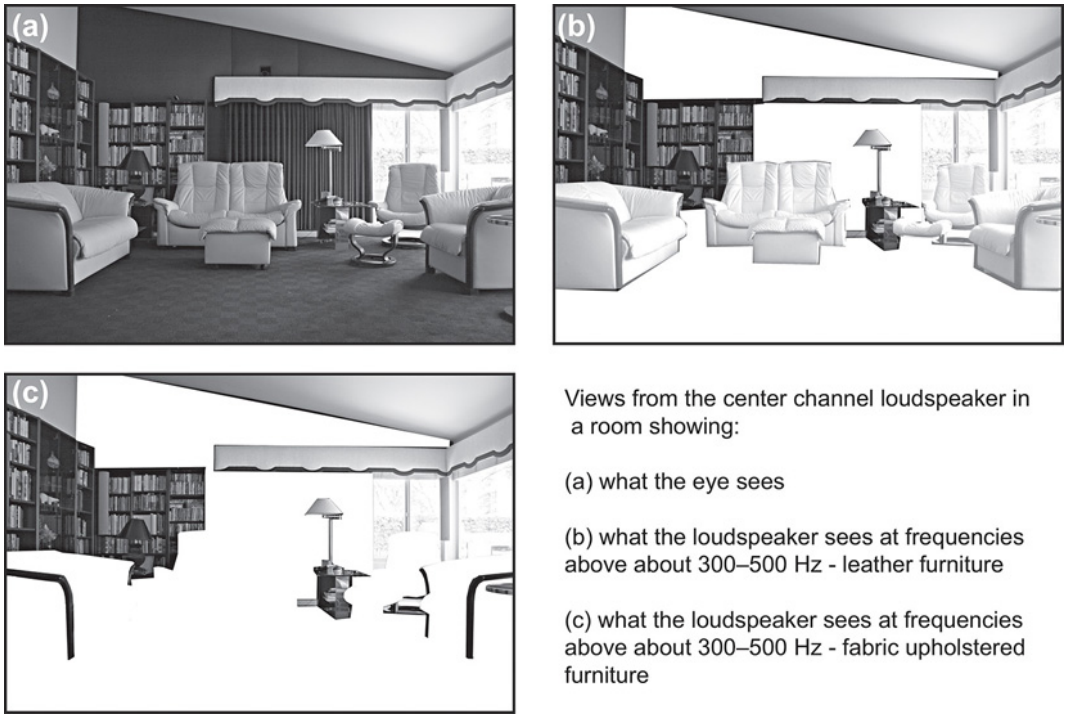


FIGURE 12.3 *A view of a listening space as "seen" by the center channel loudspeaker. It shows that much of the direct sound from the loudspeaker is absorbed at the first encounter with a surface, therefore it is unavailable to contribute to later reflections. In (b) the leather-covered furniture is shown as being lighter, because there will be some mid- to high-frequency reflection from the surface. With fabric upholstery (c), even this is gone. There is little reflecting or scattering surface area left, although a wider-angle lens would have shown that there is no side wall on the left side, it is an opening to the rest of the house, and on the right this lost energy is balanced by velour drapes over a window wall.*

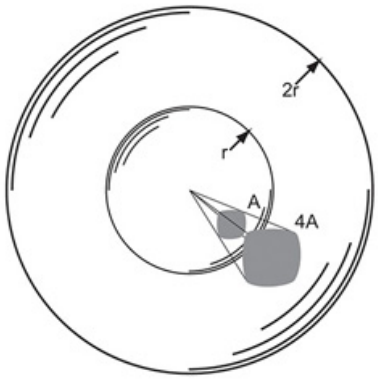
12.2 SOUND PROPAGATION BASICS

12.2.1 Point Sources

Academic discussions of sound propagation often begin with a “point source”. This is not the conventional forward-firing loudspeaker many people think of, but a theoretical sound source radiating equal energy in all directions and having no physical size. Sound from such a source is truly omnidirectional. If it radiates into free space, no reflections, the sound pressure decreases by 6 dB for each doubling of distance: this is the “inverse-square law”. In rooms this relationship therefore applies only to the direct sound. Figure 12.4 illustrates this as well as some real-world variances from theoretical ideals.

Figure 12.4 illustrates that there are complications in real-world measuring and listening. We will dig deeper into the effects of room boundaries later, but first let us get an idea of where the far field begins for conventional forward-firing loudspeakers.

(a) The Theoretical Ideal: an infinitely small sound source

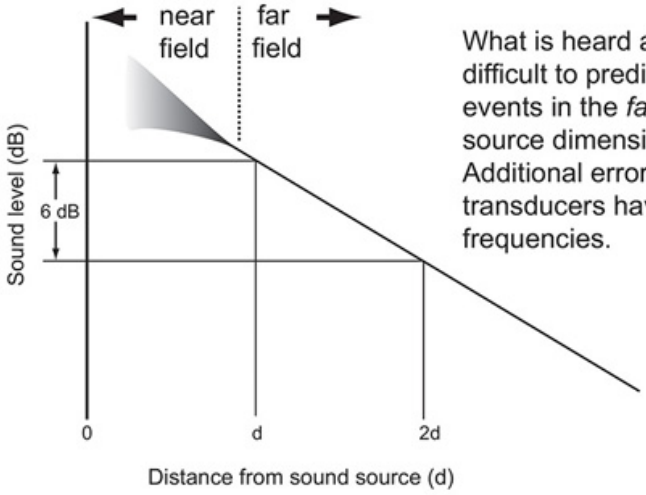


Point source / spherical spreading:

Area of spherical surface = $4\pi r^2$

When a source is small compared to the measurement distance the sound level falls 6 dB per doubling of distance (6 dB/dd). This is called the far field.

(b) The Real World: sound sources have dimensions



What is heard and measured in the *near field* is difficult to predict and not a reliable indicator of events in the *far field*. The errors depend on the source dimensions.
Additional errors arise because loudspeaker transducers have different directivities at different frequencies.

(c) The Real World: Rooms have boundaries: this is only a floor.

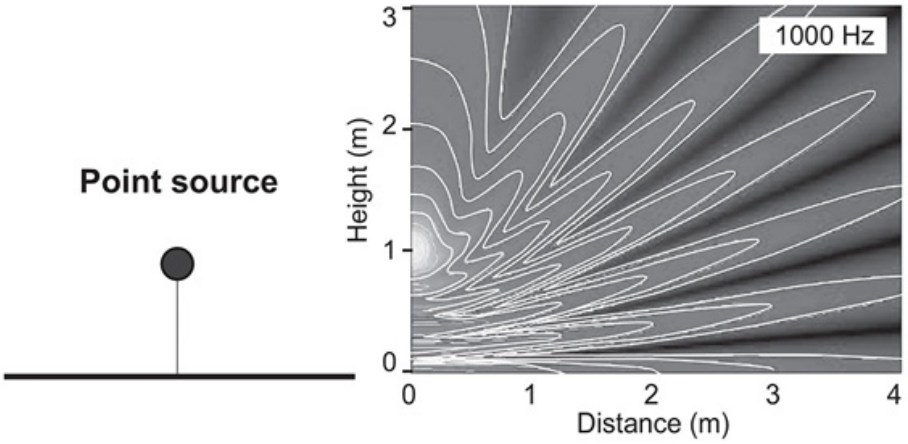


FIGURE 12.4 *(a) illustrates the theoretical ideal point source radiating spherical sound waves that spread the energy uniformly*

over an ever increasing surface area with increasing distance. This is the origin of the inverse-square law. (b) illustrates that in the real world sound sources have dimensions, and this results in measurement errors and unpredictable sound quality in the "near field". (c) illustrates a computer simulation of the effect of a single room boundary on a point source, here the floor, on the sound radiation pattern at various distances and elevations. Adjacent contour lines represent sound levels that differ by 3 dB. This illustration is for 1 kHz and it will be different at different frequencies, thus anticipating a very complex situation in the far field of a sound source in a real room. This diagram is from Keele and Button (2005). ⏎

12.2.2 Near and Far Fields Influence How We Do Measurements

When sound radiates from a complex source like a loudspeaker what is measured and heard at different frequencies is different at short distances and long distances. Within what is called the *near field* the sound measured at any point is uncertain if multiple parts or portions of the source are simultaneously radiating at the same frequency. Acoustical interference between and among sounds from the radiating elements cause position-dependent irregularities, and both microphone angular orientation and distance become variables in what is measured. In the case of loudspeakers, we need directivity information, but would like it to be independent of microphone distance. Clearly the size of the "source" is a key variable, but in loudspeakers the source size varies with frequency.

At low frequencies only the woofer(s) is active. To this might be added output from a port(s) and radiation from enclosure panels. Panel radiation would be very small compared to the acoustical output from the driver(s) and port(s) which for practical purposes define source size. The total enclosure itself is not likely a factor, nor is corner diffraction at very long wavelengths.

At middle frequencies the transducer is the primary source, with diffraction effects from the enclosure edges and baffle size. Edge diffraction of 90□ off axis sound will be greater with small transducers, and will decrease as forward directivity increases towards tweeter crossover frequency. A woofer/midrange as in a bookshelf loudspeaker will behave differently from a dedicated midrange driver that, being smaller, has wider dispersion. Enclosures designed to reduce diffraction influence this. Defining where far field might begin is not simple.

At high frequencies, unbaffled tweeters can be very widely dispersing above crossover, becoming more directional at higher frequencies. Flush mounted tweeters will disperse more widely than those with waveguides aiming to match the midrange directivity at crossover. Edge diffraction therefore depends on tweeter size, mounting and enclosure curvature. Determining where far field begins is challenging.

Then, there is the problematic crossover region within which one transducer is gradually attenuated with increasing frequency and the other is progressively activated, all determined in a manner to maintain uniform acoustical output at all frequencies while both transducers are radiating. The far field therefore is greatly extended in this region, as serious acoustical interference opportunities exist. Obviously, low-slope crossovers expand the frequency range over which multiple transducers are active, adding further uncertainty to what is measured.

Beranek (1986) suggests that the far field begins at a distance of three to ten times the largest dimension of the sound source. At this distance the source is small compared to the distance, and a second criterion is normally satisfied: $\text{distance}^2 = \text{wavelength}^2/36$. Evidently, there is academic ambiguity about where far field begins.

Based on Beranek's information Figure 12.5 shows estimated distances at which far-field conditions should prevail for a loudspeaker system and for its components. This would be the minimum distance at which a microphone should be placed for measurements and some of the distances are practical and others not so much.

Being practical means making measurements, many of which will be within the near field of the source.

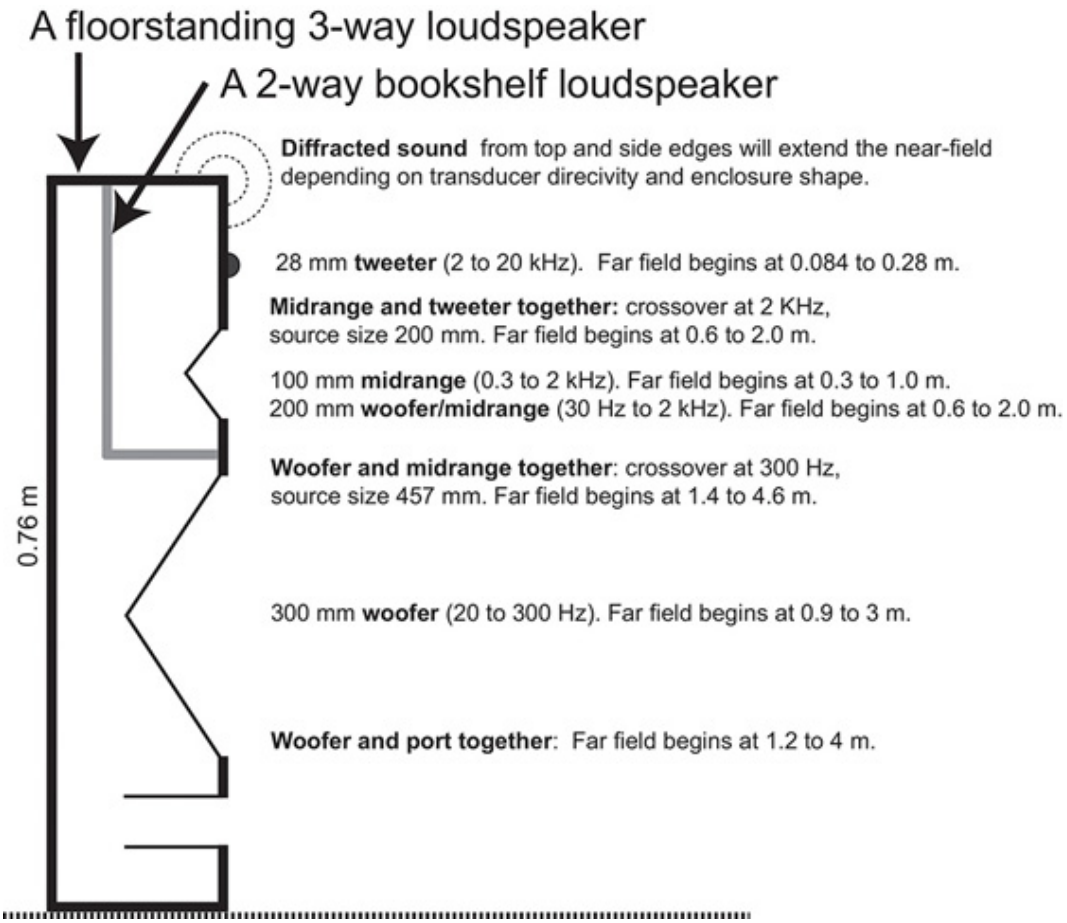


FIGURE 12.5 *Very approximate estimates of where the far-field begins for individual transducers and systems created using those transducers. This assumes an anechoic space.* ⏎

Measurement technique therefore matters. There is merit in spatial averaging, combining multiple measurements made at different angles, to minimize near-field measurement artifacts that appear in single measurements. I began spatially averaging data as soon as a digital system permitted it; see Figure 2.4. In spinorama measurements the *on-axis* frequency response is shown, but reliance is also placed on the spatially averaged *listening window* curve because it attenuates residual acoustical interference effects. All other spinorama curves are spatial averages of

multiple measurements, again attenuating evidence of acoustical interference while retaining spectral balance information and evidence of resonances.

At present there is widespread agreement that 2 m (6 ft) is a reasonable distance for making anechoic measurements to characterize the performance of conventional domestic and studio monitor loudspeakers. Large loudspeakers must be measured at greater distances, including any that involve adjacent boundaries in their function—the reflections are part of the source.

These measurements attempt to characterize the sound source in isolation, but we don't listen in anechoic spaces. "Near-field" loudspeakers are therefore often listened to within the acoustically defined near field. In fact, if viewed from the perspective of acoustical interference and point-to-point differences in sound, listeners in small rooms are all in the "near field" of the loudspeaker/room system. Only the direct sound survives relatively intact, which likely explains its perceptual importance.

This historical set of rules has been upset by a sophisticated measuring apparatus combined with significant higher mathematics and physics. The Klippel near-field scanner (NFS) scans the near sound field surrounding a loudspeaker with a robotic moving microphone, and with an understanding of the static and changing sounds is able to separate the sound radiated by the loudspeaker from that returning from the room. It then can predict acoustical events in the far field with impressive accuracy. The measurements can be done in ordinary rooms (www.klippel.de). Eliminating the need for an anechoic chamber, or the imperfections of other compromise methods, it is a significant advance in loudspeaker measurements for manufacturers and reviewers, as was discussed in Sections 2.9 and 10.2.

Diffusers behave as secondary sources of sound, and they can cover significant areas of room surfaces, so they too have near- and far-field behaviors. Cox and D'Antonio (2004, p. 37) point out that listeners should be placed as far from scattering surfaces as possible, at least three wavelengths away. This is to avoid hearing colorations from repeating

features of the devices. For devices that are effective down to 300–500 Hz, this is a minimum distance of about 10 ft (3 m). As they realistically point out: “In some situations this distance may have to be compromised.” The positioning of engineered diffusers relative to listeners is rarely a consideration, and do listeners notice?

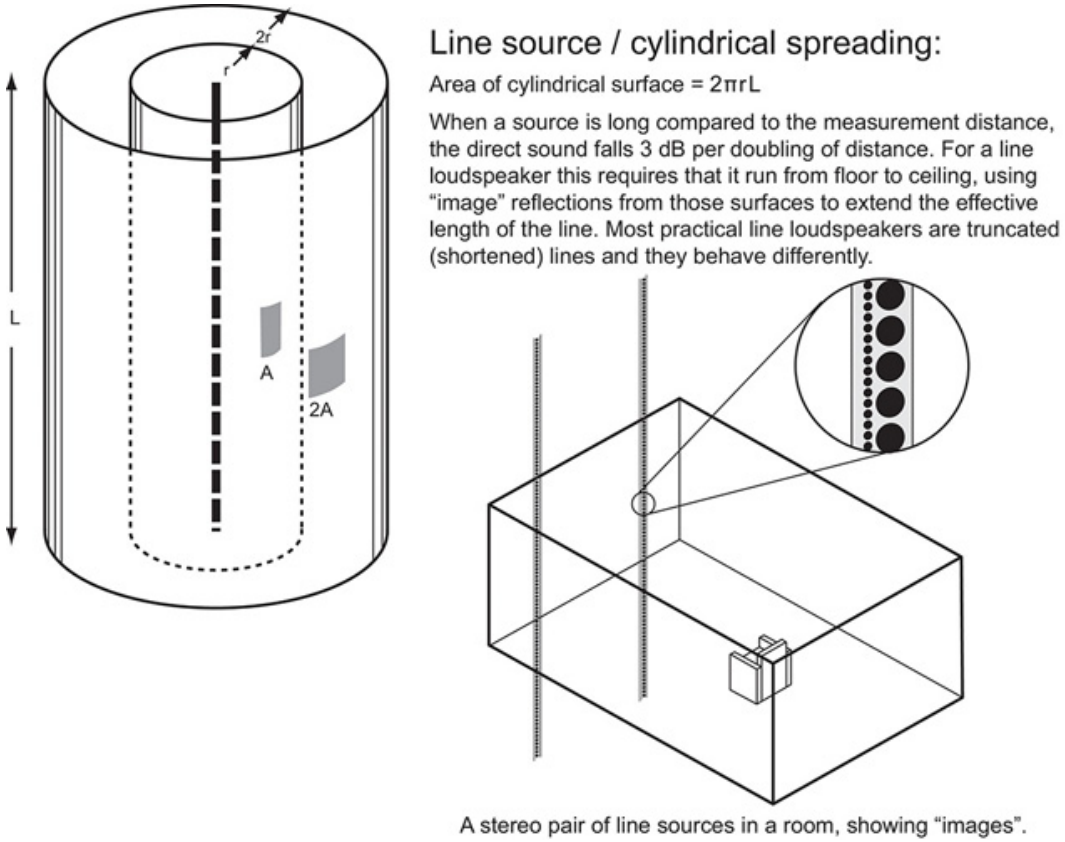
Near-field monitors. In recording control rooms it is common to place small loudspeakers on the meter bridge of the recording console or at a close listening distance. These are called near-field or close-field monitors because they are not far from the listeners. As shown in Figure 12.5 the near field of a small two-way loudspeaker (the midrange and tweeter of the example system) extends to somewhere in the range 21 inches to almost 6 ft (0.53 to 1.8 m). Including the reflection from the console under the loudspeaker, or any other nearby surface, extends that distance. Figure 12.4(c) shows what a single reflecting surface can do to what might have been a tidy radiation pattern. In fact, because the console reflection is in line with the loudspeaker, it is a classic comb-filtering situation—binaural directional discrimination is not involved. There is no doubt that the recording engineer is listening in the acoustical near field, and that what is heard will depend on where the ears are located in distance, as well as laterally and in height. This is not ideal, but it does avoid some problematic small room sounds.

Close-listening arrangements work well with the loudspeakers elevated to ear level away from reflecting surfaces. Mid-field loudspeaker locations, behind the console and arranged to minimize reflections, are excellent. In-wall mounting is an alternative for appropriately engineered loudspeakers (see Figures 13.7 and 13.8).

12.2.3 Line Sources

For a contrast, let us examine the behavior of true line sources. I say “true” because they are rare. Tall loudspeakers are common, but these are multiway systems, not a linear array of full-range transducers. Even truncated line sources behave very differently, and much worse, as will be seen. Figure 12.6 shows the orderly cylindrical spreading from an infinite line source.

Practical line sources have finite lengths, so the critical issue becomes one of keeping listeners within the near field of the line, where the desirable −3 dB/double distance propagation loss relationship holds, and out of the far field where even line sources revert to −6 dB/double-distance. Obviously, the distance at which the near-/far-field transition occurs is a function of frequency and the length of the line. So, interestingly enough, for conventional loudspeakers we may wish to be in the acoustical far field, but for true line sources we need to be in the acoustical near field to appreciate the benefits.



A stereo pair of line sources in a room, showing “images”.

FIGURE 12.6 *An illustration of a theoretical infinite line source, and of a practical approximation.* ⏎

Figure 12.6 shows a stereo pair of full-height lines, taking advantage of the ceiling and floor reflected images to make them appear to be even longer. A portion of one line has been expanded to show that it is a two-way system using conventional cone and/or dome loudspeaker drivers, densely packed (ideally spaced by less than about one-half wavelength of the highest reproduced frequency) in order to simulate a continuous sound source. Obviously, planar transducers can also be used, but they too have modular repetitions.

It is possible to use less than a full-height floor-to-ceiling array if one understands the variables and how they can be traded off. Lipshitz and Vanderkooy (1986) provide a thorough theoretical background to the behavior of "finite length" (not full height), truncated, line sources and they point out a number of problems, ultimately concluding that "there is little to recommend the use of line sources as acoustic radiators". They did grant that full-height lines had potential if the −3 dB/octave tilt in the frequency response is corrected.

Griffin (2003) gives a comprehensive and comprehensible presentation of what is involved in designing practical line sources that approach the performance of full-height lines using less hardware. Smith (1997) describes a commercial realization.

12.2.4 How Various Sources Interact With a Floor

Traditional loudspeakers employ multiple transducers, each operating over a fraction of the audible frequency range. They individually interact with reflecting surfaces around them, and they exhibit varying directivity with frequency. In tower loudspeakers it is common to see multiple woofers, and more, and this has led to them being called "line sources" or "line arrays". They consist of transducers arranged in a line, but they are absolutely *not* either line sources or line arrays. They are just tall loudspeakers employing

the same principles as their shorter versions, unless they are boundary-friendly designs (Section 13.5).

A continuous narrow strip radiator, or a line array of closely spaced small loudspeakers (e.g. Figure 12.6) is a very different device with its own distinctive properties. Computer simulations now permit us to explore domains with precision previously denied us, and loudspeaker and electronics technology allow us to turn some of it into physical reality. In sound reinforcement such designs have been used for decades to control sound radiation patterns in large, reflective venues, like churches.

In small rooms it is a significant complication that adjacent reflecting surfaces must be considered. This means that the far field for the combination of a conventional loudspeaker and one or more adjacent boundary reflections (now part of the "source") can be a long distance away. What we hear and measure at normal listening distances in small rooms is subject to variations.

Figure 12.7 is a small excerpt from Keele and Button (2005) − a collection of thought-provoking predictions and measurements. In it they examine the performance of a theoretical point source compared to several variations of truncated lines: straight and curved, "shaded" (drive power reduced towards the end) and unshaded (all transducers driven equally), all standing on a reflecting surface. For comparison I have added a floor-to-ceiling true line source at the top that embraces the floor and ceiling reflections as extensions of itself.

Figure 12.7(a) shows the orderly behavior of a true line source in its near field: −3 dB per double distance at all frequencies.

In Figure 12.7(b) can be seen how a single reflecting surface, the floor, disrupts the radiation pattern of an omnidirectional point source. Instead of tidy expanding circular contour plots illustrated in Figure 12.4(a) we see evidence of substantial acoustical interference between direct and reflected sounds generating alternating lobes of high and low sound levels heading in many directions. The constant directivity of the source, indicated on the right, means that this problem exists at all frequencies, but the patterns will

be different because of differing wavelengths. Additional boundaries, ceiling, side walls, add more of the same, of course.

Typical forward-firing loudspeakers are omnidirectional only at low frequencies, becoming more directional at higher frequencies, which helps in that the highest frequencies arrive at listeners mainly as direct sound only, negligible reflections (e.g. Figure 2.6). The sounds arriving at a listening location in a room are therefore complicated, but if the reflected sounds have a spectrum that is similar to the direct sound, the merged combination usually ends up being perceptually more satisfactory than this single-frequency, single-dimensional perspective suggests.

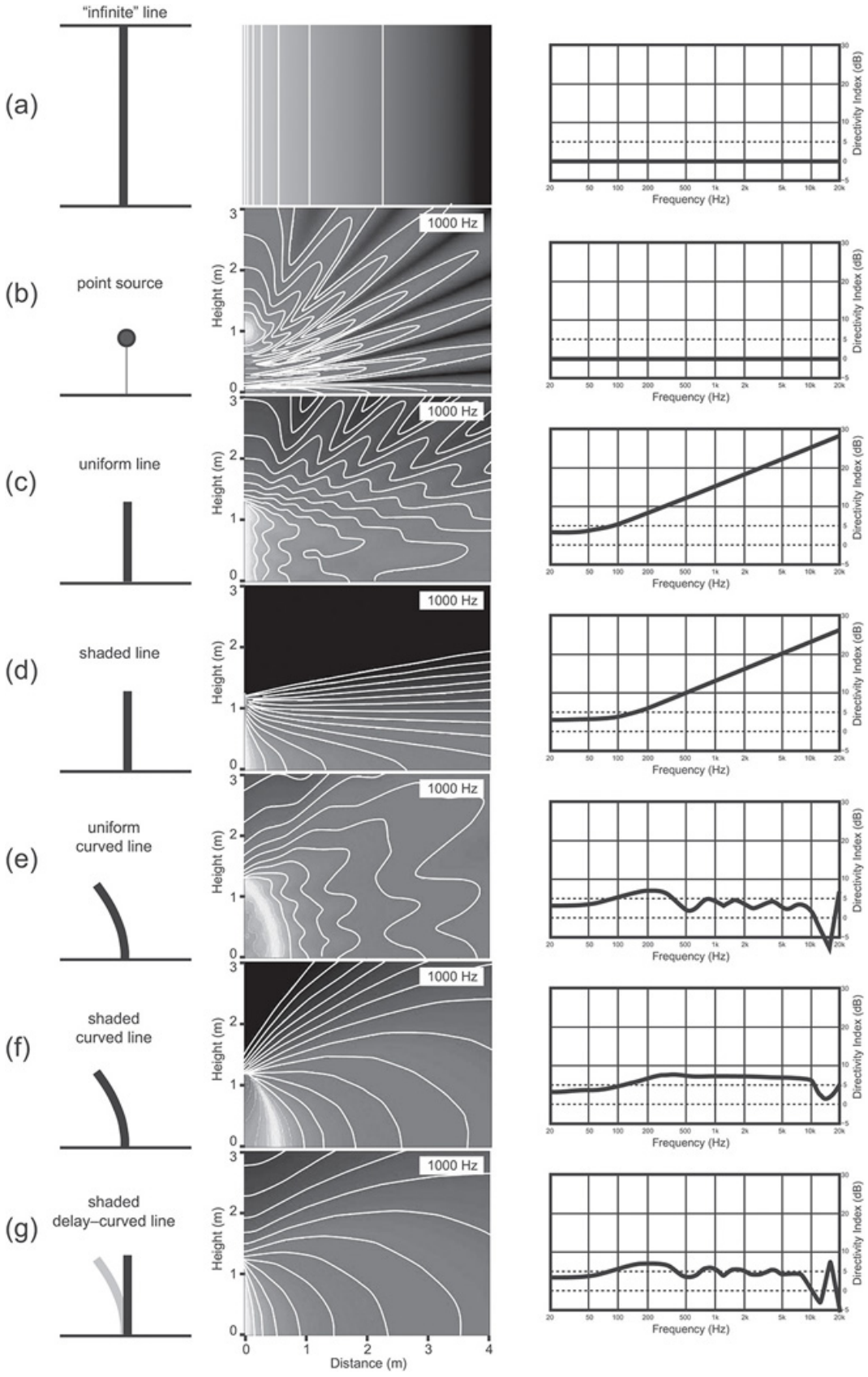

(a)
"infinite" line
(b)
point source
(c)
uniform line
(d)
shaded line
(e)
uniform
curved line
(f)
shaded
curved line
(g)
shaded
delay–curved line
1000 Hz
Height (m)
Distance (m)
Directivity Index (dB)
Frequency (Hz)

FIGURE 12.7 *Illustrations of the near sound fields generated above a ground plane by several sound sources. The author has added an idealized line source at the top as a contrast to the truncated lines shown below it. The shading in the middle illustrations gets darker as sound levels drop; adjacent contour lines represent sound levels that differ by 3 dB. The original paper displays results for several frequencies, all of those shown are for 1 kHz. The words and graphics on the left explain the sources. On the right are far-field directivity indexes. Data from Keele and Button (2005).* ⏎

Figure 12.7(c) and those that follow show what happens when considering the floor boundary as an integral part of vertical array loudspeaker designs. Figure 12.7(c) shows that a simple truncated line is an improvement over the elevated point source, but uniform directivity has been sacrificed. The directivity index has a sharply rising character, indicating high-frequency beaming.

Figure 12.7(d) shows that shading the output, reducing the drive delivered to the transducers closer to the top of the line according to a Hann contour, greatly simplifies the pattern, yielding stable sound levels at ear level over the length of the room. However, it still beams dramatically at high frequencies. We are not there yet.

Curving the line, as shown in Figure 12.7(e) is a step in the right direction. The contour lines are not yet smooth, but there is an underlying desirable order to them. The constancy of the directivity index tells us that it applies over a wide bandwidth.

In Figure 12.7(f) we see that shading the curved line using the Legendre contour yields a set of plots that have two desirable characteristics: a stable sound level at ear level out to 3 m or more from the loudspeaker, and a relatively constant directivity, meaning that it applies over a wide frequency range. This is a good result, and it has been called the Constant Beamwidth Transducer (CBT). This is not merely coping with the floor as an adjacent boundary, a low-frequency concern; it is using the floor reflection to create

a sound field that has desirably uniform broadband sound level over a wide range of listening distances.

Curved loudspeakers may not be to everyone's liking, so applying the appropriate delays to the individual drivers can, in effect, acoustically contour a physically straight line. When Legendre shading is applied to the delay-curved line the resulting CBT is very similar, as seen in Figure 12.7(g).

A comparable effect can be created without a floor, but the array will need to be differently optimized. JBL Professional has a series of innovative free-standing CBT products that have found use in sound reinforcement as well as high-end cinema and home theater surround applications. The stability of both sound quality and sound level as a function of listening distance is striking, which makes such designs generally attractive and especially for use as side surrounds (see Section 8.5.4). I have experienced several CBT designs and they all deliver on the promise. One is able to start at a large distance, walk towards the loudspeaker, finally stand adjacent to it, and the sound level hardly changes. It defies instinct.

12.2.5 Are Floor Reflections Special?

Very early in my explorations of loudspeaker/room interactions I was aware of the measurable effect of the floor reflection in steady-state room curves. It seemed like a problem that needed attention, so I devoted some time to modifying loudspeakers to minimize it, and followed through with subjective evaluations. I cannot claim to have been exhaustive, but I soon became frustrated when the curves looked better but the sound seemed not to have changed very much. I moved on to other problems that were clearly audible and never returned.

Siegfried Linkwitz also decided that "the floor reflection, which is mostly unavoidable, and which is readily seen in the steady-state on-axis frequency response as dips and peaks, is not necessarily audible on program material" (Linkwitz, 2009).

The Fraunhofer Institute in Germany constructed an elaborate listening room in which different room surfaces could be changed (Silzle et al.,

2009). “Regarding the floor reflection, the audible influence by removing this with absorbers around the listener is negative—unnatural sounding. No normal room has an absorbent floor. The human brain seems to be used to this.”

There is an intriguing logic to the notion that humans evolved while standing on something, and most of what we hear includes reflections from what is under our feet. Wherever we are, reflections from what is under our feet have been useful in gauging distance, among other things. Floor reflections are ever-present. Is that a problem, a virtue or just a fact of life?

12.3 PROPAGATION IN ROOMS WITH AND WITHOUT REFLECTIONS

12.3.1 Classic Concert Hall Acoustics

A discussion on acoustics often begins with explanations of events in reverberant spaces, a concert hall being the most thoroughly studied of all. These spaces have the difficult task of preserving the finite amount of energy radiated by musical instruments and voices (meaning as little absorbing material as possible), delivering it to all parts of a large audience (meaning large carefully angled and shaped reflecting surfaces), while adding the right amount of reverberation for the kind of music being performed (meaning high ceilings, large volumes, to allow the sound to reflect multiple times). It is not simple.

Figure 12.8 shows the textbook explanation of concert hall performance, where the direct sound is seen to fall according to the inverse-square law (−6 dB per double distance) until it meets the reverberant sound field. The distance at which they are equal is called the critical distance, or the reverberation radius or distance. It is the point where a draw-away measurement of sound level transitions to remaining relatively constant. In many concert halls this is only a few rows into the audience.

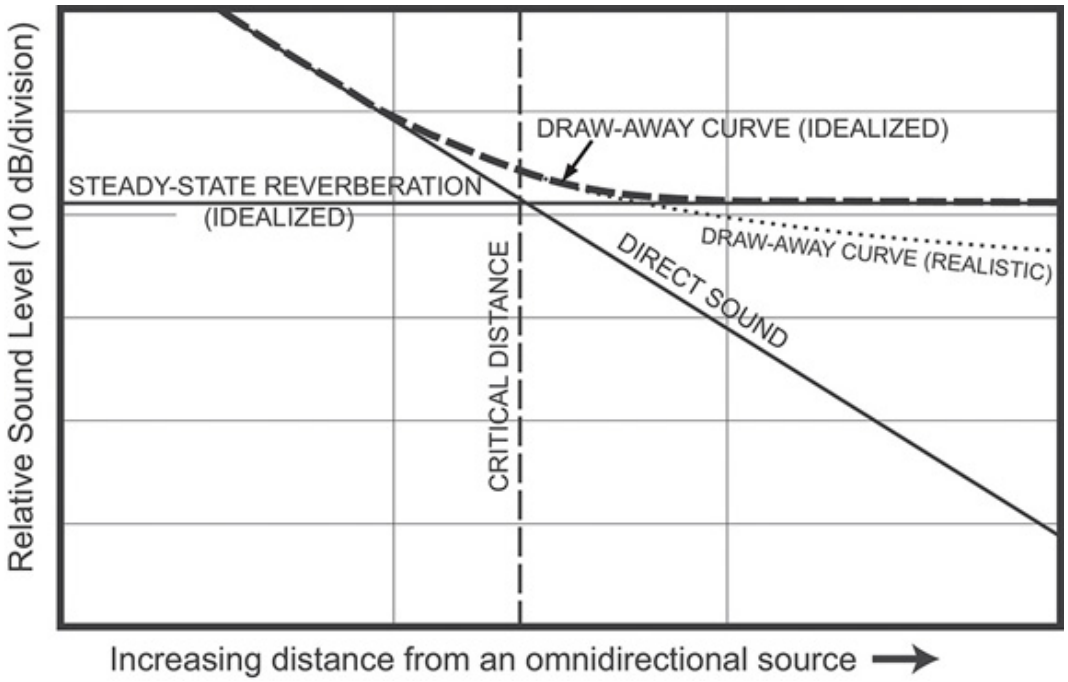


FIGURE 12.8 *The structure of the steady-state sound field in an idealized large performance space. The dotted draw-away curve indicates the declining sound level that, in some form or other, is seen in real halls. Based on Schultz (1983, figure 10).* ⏎

If the total amount of absorption in the hall is increased, RT drops, and the steady-state reverberation is lowered in level. The critical distance moves farther into the audience. If the directivity of the sound source is increased, less energy is delivered to the reverberant sound field and again the critical distance moves farther from the source. In sound reproduction, loudspeakers have different directivities for different applications. Figure 12.9 shows directivity indexes for loudspeakers and other relevant sound sources.

Some musical instruments, especially stringed instruments, exhibit very complicated radiation patterns, so these are simplifications; Meyer (2009) shows much more detailed data (e.g. figure 4.1). It is important to note that they fall into a similar range of directivities as the loudspeakers. Horns in both instances are most directional. Very clearly shown in the loudspeaker data is the importance of the size of the low-frequency energy source, with

the directivity index curve progressively flattening as the radiating area—the size and number of woofers—shrinks. There is also a reduction in DI as the angular dispersion of the high-frequency horn expands, reducing further with the small dome tweeter in the domestic loudspeaker. Obviously, the more directional loudspeakers were designed to deliver strong direct sounds to listeners at larger distances.

Loudspeakers designed with flat on-axis frequency responses, so as to accurately reproduce the initial timbral signature of the recorded sounds, will therefore exhibit a rising sound power output at lower frequencies. This results in a downward-tilted steady-state room curve, which some people find puzzling. This is normal, and familiar to humans, because it is true for voices and musical instruments as well—in fact, all sound sources that are small compared to wavelength at low frequencies; see Figure 9.16. Whatever *direct* sound spectrum is radiated by the source, there will be a relative rise in the bass *sound power* that is radiated.

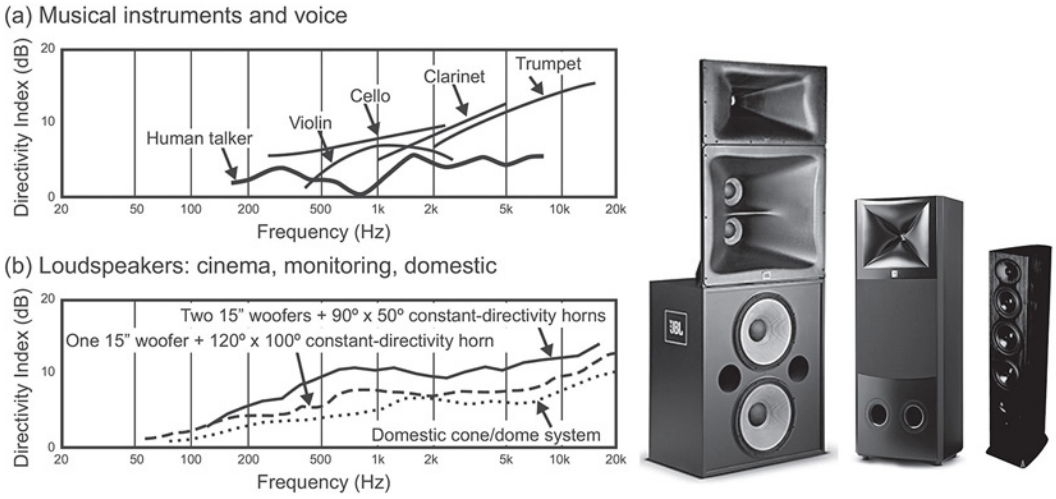


FIGURE 12.9 *(a) The simplified directivity indices (DI) for some musical instruments (Meyer, 2009). The voice data are from the first edition of this book, figure 10.3. (b) Directivity indices for a large cinema loudspeaker system with a double-woofer bass unit and 90° × 50° horns (JBL Pro 5732), a single woofer studio monitor system with a 120° × 100° horn (JBL Pro M2), and a*

three-way domestic cone/dome system (Revel F206) which are shown approximately to scale. Based on Toole (2015, figure 2). ⏎

In real terms these data tell us that during a concert the critical distance varies, being longer for directional horns than for multidirectional violins.

In a measured steady-state room curves for loudspeakers one can anticipate that the bass will be elevated compared to the direct-sound, on-axis curve. The amount of the bass rise will depend on the reflectivity of the room at low frequencies, which is represented by the RT of the room. Figure 12.1 shows that the variations in bass can be considerable. The increasing directivity with frequency also means that listeners will be in a progressively more direct-sound dominated sound field at high frequencies.

In some publications the region ahead of the critical distance is called the "near field" and that on the distant side is called the "far field". This is an unfortunate use of the terms because, as has been seen in Section 12.2, there is another use related to sound sources with, I think, a longer history. More accurate terms could be "direct-sound dominant" and "reflected-sound dominant" describing circumstances in the performance venue.

12.3.2 Sound Propagation in Typical Listening Rooms

The situation is very different in the small, more absorptive, rooms used for sound reproduction. Loudspeakers have significant directivity, gradually increasing towards high frequencies. This means that the critical distance would be frequency dependent. In addition, the reverberation times are so short that the concept of a "steady-state" background reverberation is dubious, and that is the basis for the critical distance.

Chapter 4 in the first and second editions of this book explained a progression of acoustical changes starting with concert halls, moving through industrial and commercial spaces, and ending with those that pertain in sound reproduction venues. The intermediate phase is also shown in figure 3 of Toole (2006). Part of the difference is attributable to the amount of absorbing material and where it is deployed. Some of the difference is related to the low ceilings and the presence of surfaces and

proportionately large objects—including humans—that reflect some sound back towards the source. The end result is summarized in Figure 12.10.

Figure 12.10 shows the manner in which the direct sound and the direct sound combined with reflected sounds behave in typical domestic listening spaces. The A-weighted measurement ensured that bass standing waves were not a factor (Figure 10.17). In these small, low reverberation rooms, there will be no truly diffuse "reverberation". The reflected sounds contributing to the elevated steady-state sounds are best described as early reflections, some of which are reflected back towards the source by objects and surfaces in the room. Obviously the less reflective the listening environment, the steeper will be the draw-away curve, ultimately uniting with the −6 dB/dd curve for a completely non-reflective, anechoic, environment. Close to the loudspeaker, and certainly at 1 ft (0.3 m), the microphone is in the acoustical near field of the loudspeaker, and spectral variations are likely to occur (Section 12.2.2).

Of separate interest are the data points from the Zacharov et al. (1998) study, which were determined not by measurement but by subjective loudness matching. They fall on the −2.5 dB/dd line, very likely decided by the specific combination of loudspeaker directivity and room acoustics used in the experiments. This is reassuring in that we now have confirmation that these graphic data can be used to anticipate the perceived loudness of steady-state acoustical events in typical rooms.

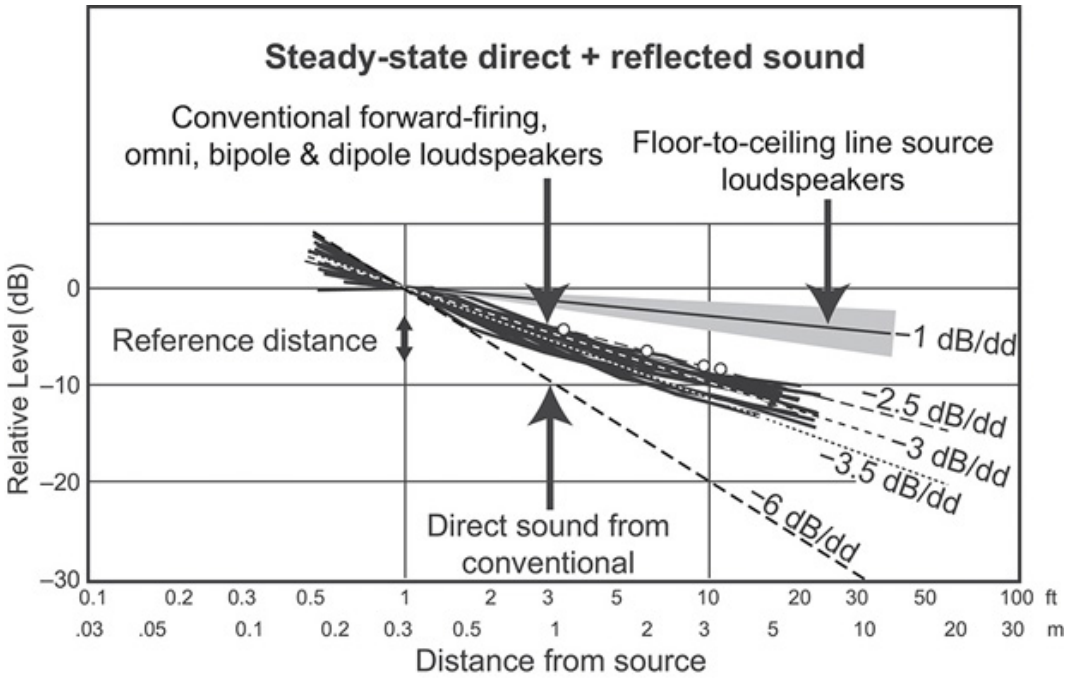


FIGURE 12.10 *The cluster of curves consists of draw-away measurements of steady-state sound in four living rooms using four approximately omnidirectional sound sources (Schultz, 1983). dd = double distance. These are combined with measurements by the author using five loudspeakers with different directivities (omnidirectional, dipole, bipole and forward-firing) in two domestic listening rooms. All curves were measured using A-weighting of broadband sounds and were normalized to an arbitrary reference distance of 1 ft (0.3 m). Through the cluster are drawn three lines having different slopes, ranging from −2.5 dB/dd to −3.5 dB/dd, embracing all of them. A reasonable average is −3 dB/dd. The four white dots are data from Part II of the experiments discussed in Zacharov et al. (1998) in which listeners did subjective loudness matching of loudspeakers at different distances. These fall on the −2.5 dB/dd line. At the top is the estimated draw-away curve for a floor-to-ceiling line source loudspeaker. At the bottom is the inverse-square law, −6 dB/dd attenuation of the direct sound with distance.* ⏎

At the top of Figure 12.10 the anticipated draw-away curves for a full-height line source are shown. The upper direct + reflected sound curve hovers around −1 dB/dd. It is based on very limited data but seems reasonable.

The sound field in small listening rooms is not diffuse, statistical acoustics do not apply, and there is *no critical distance*, as classically defined.

12.3.3 The Meaning of Steady-State Room Curves

In the preceding discussions we have considered the total steady-state sound level at various distances from loudspeakers in small rooms. But, what exactly is the sound that is being measured? The most common description is a steady-state frequency response: a room curve.

When a steady-state frequency response room curve is measured a curve results, but what does it mean? It is audacious to think that this one curve represents the complex timbral and spatial perception that two ears and a brain perceive from the combination of direct and reflected sounds. An evaluation of sound quality might be the reason for the measurement, but listeners extract much more information from the combination of arriving sounds, spaciousness being a major factor—we hear “the room”, as well as the loudspeaker. In multiple-loudspeaker listening tests we know that listeners are able to perceptually separate the changing timbral characters of the loudspeakers from the fixed sound of the listening room. They recognize subtle performance attributes in the loudspeakers being evaluated although the room sounds are present in them all. But when there are no comparisons, as in a single-stimulus listening evaluation, such an analysis is not possible (Section 1.4). The same is true with a room curve. The measurement microphone simply adds the direct and all reflected sounds together and gives us the sum. Analysis is not possible, but analysis is necessary.

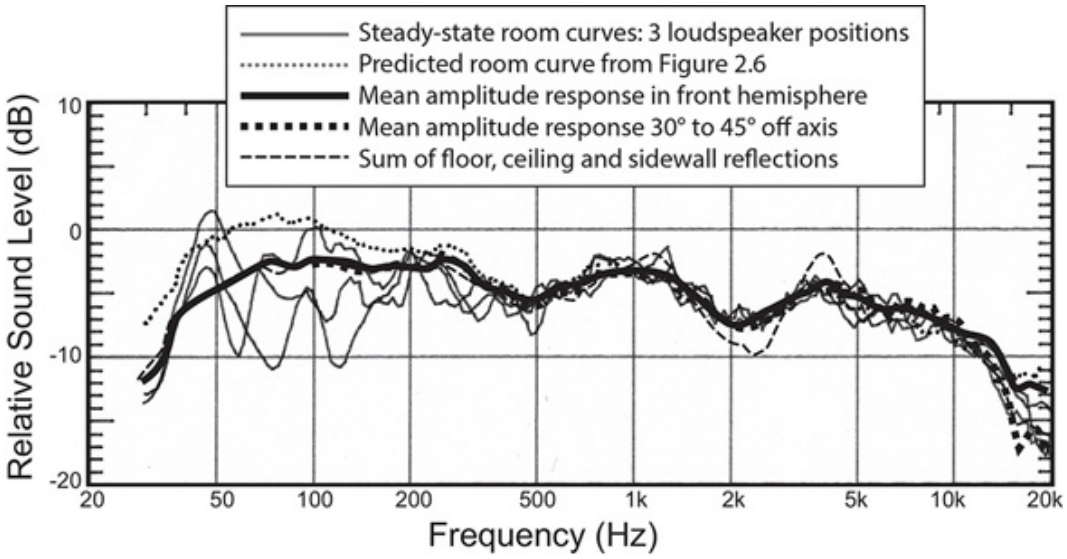


FIGURE 12.11 *Steady-state room curves measured for three different loudspeaker positions are compared to various possible predictive methods. Data from Toole (1986).* ⏎

Room curves have a long tradition in audio, and when I started my research, it was one of the logical measures to be evaluated. The predictive experiment done in 1986, shown in Figure 2.6 showed that logically selected anechoic data could be combined to predict a steady-state room curve above the transition frequency. This led to other ways to estimate a room curve, the results of which are shown in Figure 12.11. All this was to understand which sounds radiated by a loudspeaker were likely to be responsible for what listeners perceive.

Figure 12.11 shows the expected highly variable low-frequency measurements because of different coupling to standing waves at different loudspeaker locations. It also shows room curve predictions arrive at using four different methods:

1. The method used in Figure 2.6, combining direct, early reflected and sound power estimates.
2. Combining only the 25 measurements in the front hemisphere.

3. Combining only the 8 measurements at 30° and 45° off-axis horizontally and vertically.
4. Combining only the 4 specific off-axis measurements that were thought to represent floor, ceiling and side wall reflections.

Above the transition frequency region, and certainly above 500 Hz, all predictions perform well, but method 4 exaggerates the shape, indicating that more sound components are needed. At low frequencies method 1 stands alone as being too high, indicating that sound power is not a good estimator of bass performance in real rooms. The other three methods closely agree, but none can capture the extreme variability of in-room curves.

Forty years later, the situation has not changed. Figure 12.12 shows some spinorama data comparisons with real room curves.

Figure 12.12(a) loudspeaker: Infinity Prelude MTS, ca. 2000. Top: the full anechoic spinorama. Bottom: steady-state in-room measurements with the loudspeaker in the front-left location averaged over six listener locations within a typical listening area in a typical rectangular room. As a means of understanding which sounds from the loudspeaker contribute to this measurement, the early-reflections and sound power curves from the anechoic data set at the top have been superimposed. Obviously, it is the off-axis performance of the loudspeaker that is the dominant factor in determining the sound energy at the listening location and, in the several comparisons that have been done, the early-reflections curve has been a better fit than the sound power curve, which is excessively tilted. As noted in Figure 2.6(b) the on-axis curve is the dominant factor at very high frequencies, so both of the overlaid curves underestimate the steady-state room curves at high frequencies. More direct sound is needed in the prediction. As is expected, the standing waves in the room take control at low frequencies and the prediction fails.

Figure 12.12(b) loudspeaker: B&W 802N. This loudspeaker has both frequency-response and directivity issues. In the DI curves one can see the

directivity of the woofer rising with frequency, then falling in the crossover around 350 Hz to a rather large (6-inch, 150 mm) midrange driver that exhibits increasing directivity up to around 2 kHz before crossing over to an unbaffled tweeter with wide dispersion which, by 5 kHz has taken over. In the lower box of curves the pronounced mid- to upper-frequency undulations are clearly associated with the off-axis behavior of this loudspeaker system. The shape of the room curve is clearly signaled in the shapes of both the anechoic early-reflections and sound power curves.

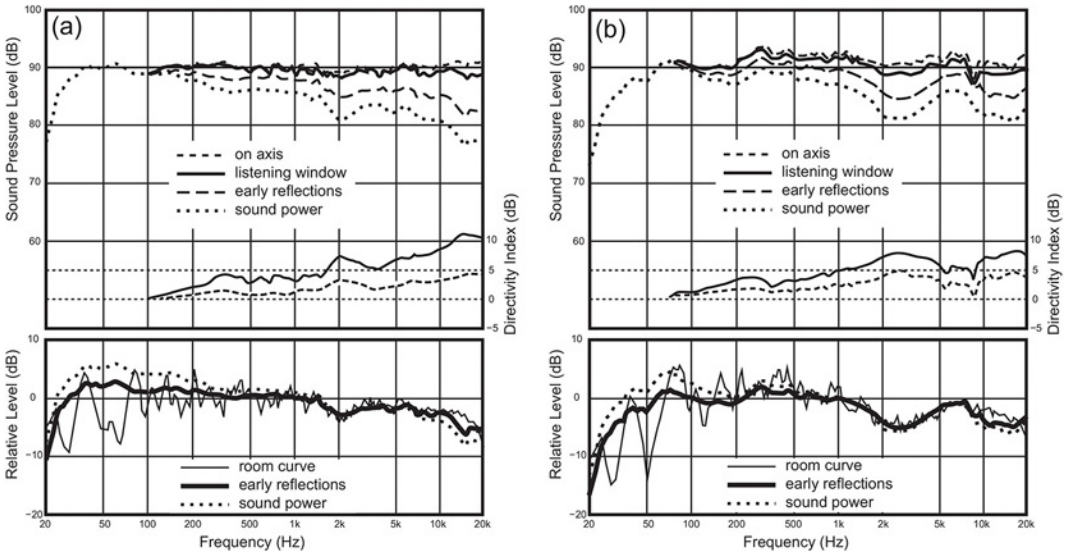


FIGURE 12.12 *A comparison of spinorama early-reflections and sound power curves to steady-state room curves for two loudspeakers.*

As indicated in Figure 2.6(d), sound power matters progressively less with rising frequency. This explains why the frontal hemisphere data, exemplified by the early-reflections curve, is a better match to the room curve. The sound power curve overestimates the low-frequency room curves and both it and the early-reflections curve underestimate the high frequencies. This is a reason to blend in at least high-frequency data from the listening window to improve the frontal weighting of the data. In no case is the on-axis frequency response by itself a useful indicator of how loudspeakers will measure in rooms—although it is critical to anticipating

how loudspeakers sound in rooms. It bears repeating: two ears and a brain are not equivalent to an omnidirectional microphone and an analyzer.

An obvious message from Figure 12.12 is that measuring a steady-state room curve and equalizing it cannot guarantee excellent sound. Equalization can only change frequency response. In the loudspeaker in Figure 12.12(b) the dominant problem is non-uniform directivity as a function of frequency, and equalization cannot change directivity. The solution is a better loudspeaker. If the only information available is a room curve, this fact is not known. Comprehensive anechoic data on the loudspeaker would have revealed this in advance.

Finally, another approach to predicting room curves was proposed by Devantier (2002), where, there is a prediction based on a combination of three spinorama curves. This process is also included in ANSI/CTA-2034-A (2015). Figure 12.13 shows the loudspeaker used as an example in that document, which states:

> It [the prediction] shall be comprised of a weighted average of 12% Listening Window, 44%, Early Reflections, and 44% Sound Power. The sound pressure levels shall be converted to squared pressure values prior to the weighting and summation. After the weightings have been applied and the squared pressure values summed they shall be converted back to sound pressure levels.

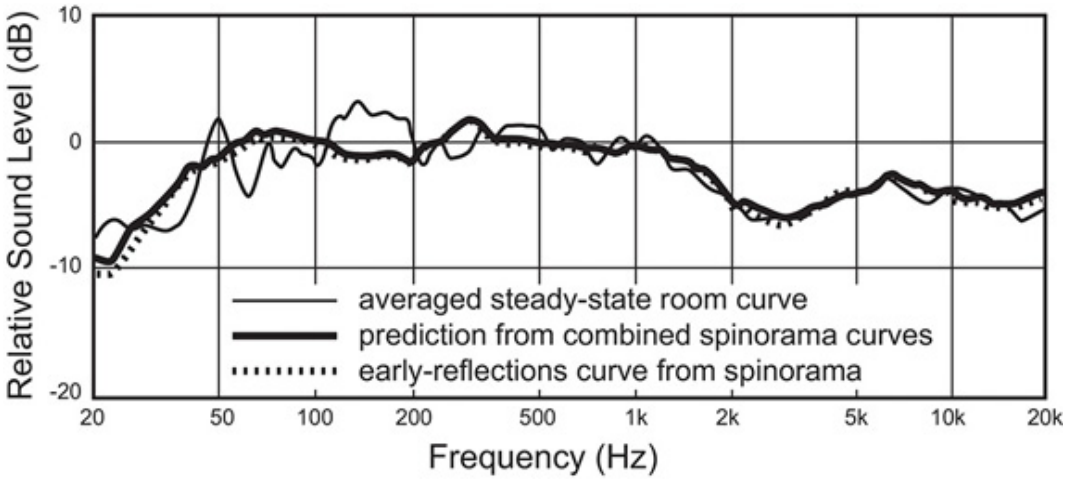


FIGURE 12.13 *Measured and predicted room curves. One curve (heavy solid line) is based on a combination of 12%*

listening window, 44% early reflections and 44% sound power. The other curve (dotted) is the early-reflections curve from the spinorama. ⏎

As Figure 12.13 shows, the agreement between the calculated prediction and the spatially averaged steady-state room curve is good above the transition frequency. It is logical to add on-axis information to elevate the very high frequencies, but with access to the original spinorama data it can be seen that, for this loudspeaker in this room, the early-reflections curve by itself is a similarly good predictor over most of the frequency range.

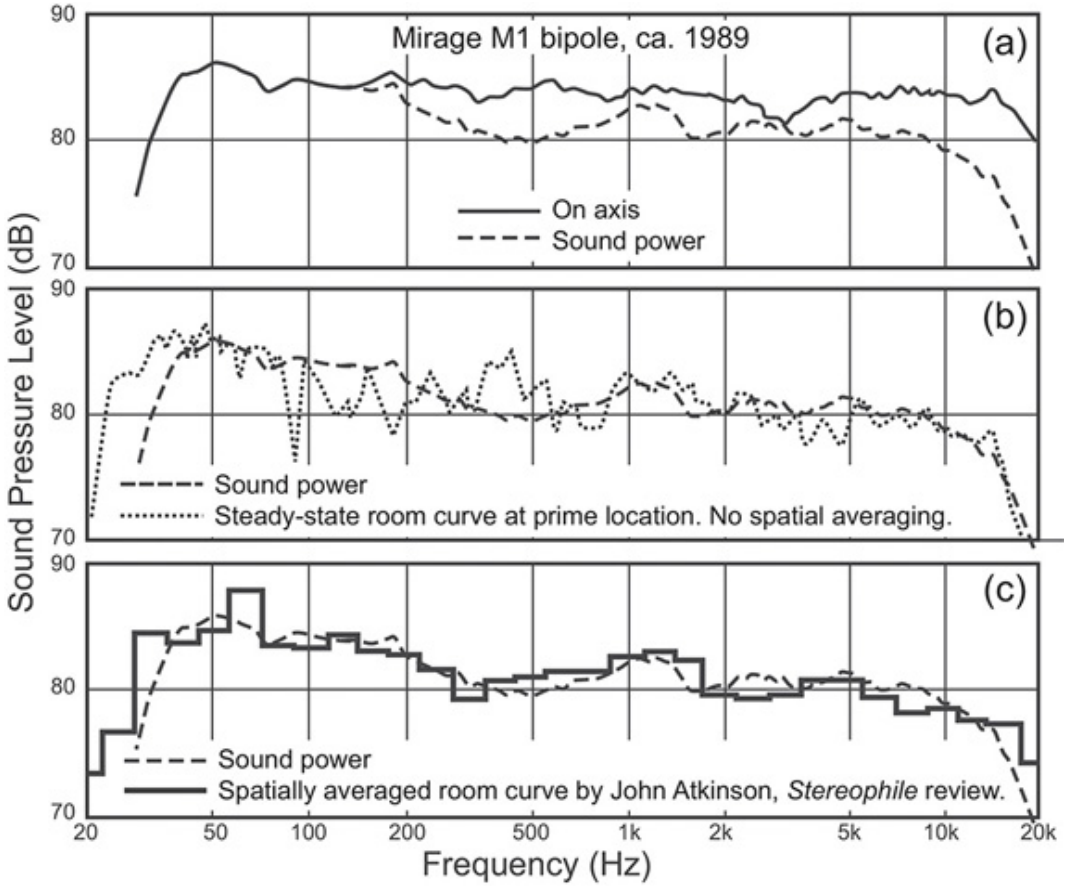


FIGURE 12.14 *(a) Anechoic on-axis and sound power measurements on the bipole (bidirectional in phase) radiating Mirage M1. (b) shows a comparison between the steady state room curve in my listening room and the sound power. (c) shows a 1/3-octave band room curve measured by John Atkinson for a*

review published in Stereophile magazine (June 1989). Used with permission. ⏎

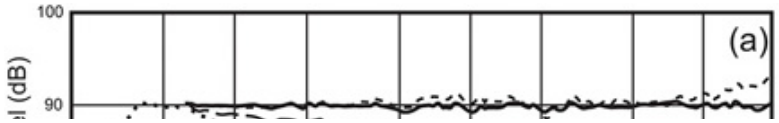
These predictions apply well to conventional forward-firing loudspeakers but there are other designs including bipole and omnidirectional loudspeakers. I had experience with a good example of a bipole, bidirectional in-phase design that got high sound quality ratings.

Figure 12.14 shows anechoic measurements and steady-state room curves in my listening room, and in another that appeared as part of a review.

Measurements shown in Figure 12.14 demonstrate that for loudspeakers approaching omnidirectionality it is sound power that predicts the shape of the room curve. Loudspeaker directivity is the determining factor for room curves, and on-axis/listening window curves are those most related to perceived sound quality.

12.3.4 One Loudspeaker in Many Rooms

All the loudspeakers examined so far have been cone/dome forward-firing designs. It is in their nature to exhibit directivity that changes with frequency. Figure 9.15 showed performance data on a loudspeaker with very constant directivity horn above the 800 Hz crossover frequency, a JBL Professional M2, a monitor loudspeaker consisting of a 15-inch (381 mm) woofer and an compression driver and horn. It is an active loudspeaker with dedicated outboard electronics, equalized to be relatively flat in the listening window. Designed for control-room monitoring and home theater applications it is capable of generating entertaining sound levels in cinema sized venues. As part of a study on cinema sound this loudspeaker was set up in front of the screen, at the center location, in six film sound venues. Measurements were made at many seating locations at ear heights and at differing microphone heights. There was no room equalization (Gedemer, 2015). Figure 12.15 shows some of the results.



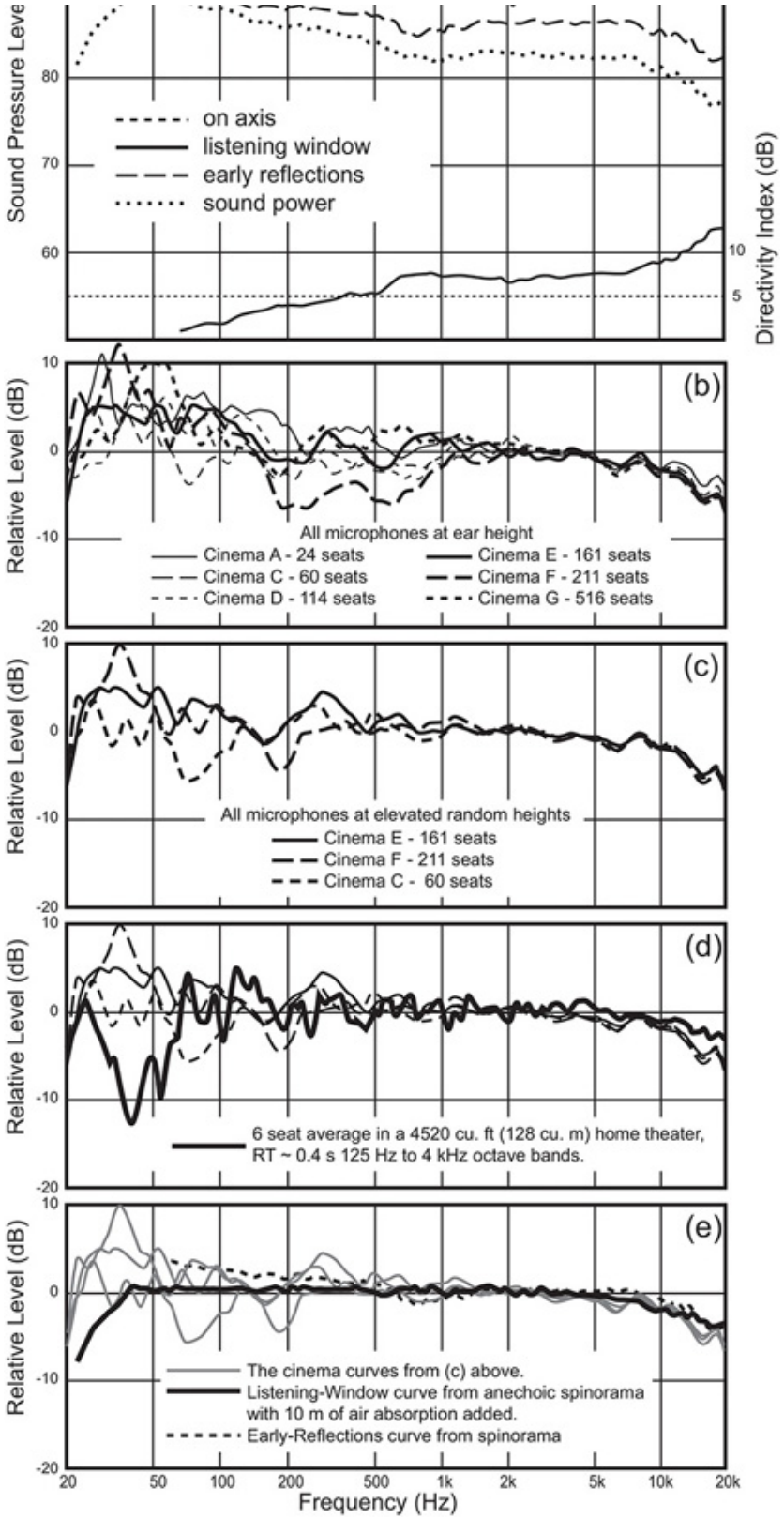


FIGURE 12.15 *The spinorama of a JBL Pro M2 loudspeaker is shown in (a). The loudspeaker was set up to radiate directly into*

six different film-sound venues. (b) shows the result of microphones at nominal ear height. (c) shows the results when three of the cinemas were re-measured with the microphones at randomly changed elevated heights. (d) shows the steady-state cinema measurements from (b) with a six-seat average of the same loudspeaker measured in a typical domestic home theater. All cinema data are shown with 1/6-octave frequency resolution. The home theater curve has 1/12 octave resolution and therefore is more irregular. (d) superimposes the listening window curve from the spinorama in (a) with 10 m of air absorption included because of the long listening distance in these cinemas—the measurement was done at 2 m. The unadulterated early-reflections curve is also shown. Data from Gedemer (2015) with additions in (e). ⏎

For reference, Figure 12.15(a) shows the spinorama on the JBL Pro M2 used in the following tests.

The curves shown in Figure 12.15(b) are averages of steady-state measurements made at every second seat in each of the facilities, except for cinema G where 109 locations were used. Above about 1 kHz the curves agree very well, indicating minimal room effects. The curves were normalized to this frequency range. Below about 150 Hz the variations are of a kind expected for differing amounts of reflected energy, adjacent boundary effects and, in the smaller venues, room modes. The irregular behavior in the 150 to 800 Hz range was troubling. Seat dip or other seat interaction effects were suspected (the microphones were at estimated ear level: 40 to 47 inches [1.0 to 1.2 m] above the floor).

Ear level microphones in close proximity to orderly rows of seats, some of which have high backs, generate effects more complex and of wider bandwidth than the well-known seat-dip phenomenon in concert halls. The fact that the cinema sound field is considerably less reflective is a probable factor. It is unlikely that these non-minimum-phase acoustical interference effects can be treated by equalization, if indeed they need correction at all.

Two ears and a brain are much more analytical and adaptable than an omnidirectional microphone and analyzer. Once stationary in a chair, listeners usually are unaware of, or attribute innocent spatial descriptors to measured variations suggesting repugnant interference/comb-filtering phenomena.

Anecdotal information, personal experience, and experiments by Newell et al. (2015) indicate that a stand-up/sit-down comparison yields an insignificant change in sound quality. If so, delivery of a high-quality direct sound can be ascertained using microphones elevated above the seats thereby avoiding distracting measurement anomalies, especially if these irregularities lead to "corrective" equalization, which will degrade the important direct sound.

Figure 12.15(c) shows that when the microphones were elevated to heights randomized over the range 51 to 79 inches (1.3 to 2 m) in three of the venues the curves congregate more closely to each other above about 400 Hz. Newell et al. (2015) suggest a distance of at least 2 ft (60 cm) above the back of the seat to avoid acoustical interference. It seems reasonable.

Figure 12.15(d) drops the audience size to a low of 6 seats in a much smaller domestic home theater. The superimposed curve shows excellent agreement but is slightly elevated at high frequencies because of the reduced listening distance (less air absorption). Except for the lowest frequencies, the cinema curves and the home theater curve appear to fluctuate around the same central tendency.

Figure 12.15(e) superimposes the listening window curve from the spinorama in (a). Above about 500 Hz listeners in cinemas are in a progressively more dominant direct sound field so this should be a good predictor of what is measured. With the addition of some air attenuation, it is. At low frequencies there is only slight bass rise, as is expected in venues that have been as thoroughly acoustically treated as these professional screening rooms have been (Figure 12.1(d)). For comparison the early-reflections curve is also shown. It anticipates the gently rising low bass, but the small 800 Hz crossover dip is seen in only one of the room curves,

possibly because of the dominance of the direct sound in these large well-damped screening rooms.

Summarizing: In all of the measurements, above about 500 Hz the steady-state room curve would be well predicted from anechoic loudspeaker data. Gedemer (2015) showed that in a 516-seat cinema, there was very little difference between the spatial averages of four microphone locations and 109 locations. At low frequencies, standing waves and adjacent boundary effects dominate results and equalization is a remedy for some of these.

Even with seat counts that range through 6, 24, 60, 114, 161, 211 and 516, there is no obvious pattern of change in steady-state room curves above about 400 Hz except the small difference in high-frequency air attenuation in the direct sound related to listening distance. Air attenuation can be calculated for any seat in any venue (see Section 12.3.6). All curves in all venues reflect expectations from anechoic data except at low frequencies where room resonances and boundary interactions dominate. Only at low frequencies are room curves revealing new information, and a basis for equalization or other forms of room mode control.

If for any reason broadband steady-state room curves are measured in cinemas, microphones should be elevated well above the seat backs, preferably at varied heights to average out regular patterns of interference.

12.3.5 The "Harman" Room Curve Is a Result, Not a Target

There have been multiple "Harman curves" over the years. They have been named by others, not by Harman, and their meaning and use have been routinely misconstrued. Going back many years, JBL Synthesis systems, upmarket home theater packages, included a "calibration" service to finalize the installation. This was done using in-room measurements, room curves, and there was a target curve that had a slight bass boost, because that was what sounded right to most customers for movies at their chosen playback levels—usually below calibrated cinema level. This was followed by a smooth, almost straight downward tilted curve. Most EQ adjustments were done at bass frequencies to address room resonances. How did we

know that this room curve was the right target? Because Harman designed, manufactured and measured the loudspeakers. It was known that if this curve was measured in situ, the loudspeakers were working properly and the listeners would be receiving an essentially flat direct sound, the one that is primarily responsible for perceived sound quality.

Not surprisingly, information about this curve got around, and it started to be thought of as a universal target curve for any loudspeaker in any room. People would equalize their systems to match the "Harman" curve, and then the discussions would begin, because with unknown loudspeakers the audible results were not predictable. The direct sound dominates sound quality, equalization changes the direct sound and a steady state room curve measures mostly reflected sound. It is a scheme that cannot work reliably.

More recently, another "Harman" curve has joined the internet audio forum discussions. This one originated in the previous (third) edition of this book. Again, it was named by others, and its meaning misunderstood. I said in the introduction to that section (12.2): "As the story evolves, it will be clear that the 'right' room curve results from a good sounding loudspeaker, but equalizing an inferior loudspeaker to achieve the 'right' room curve does not guarantee good sound. Puzzled? Read on." Figure 12.16 is a duplicate of figure 12.4 in the third edition, which followed that introduction.

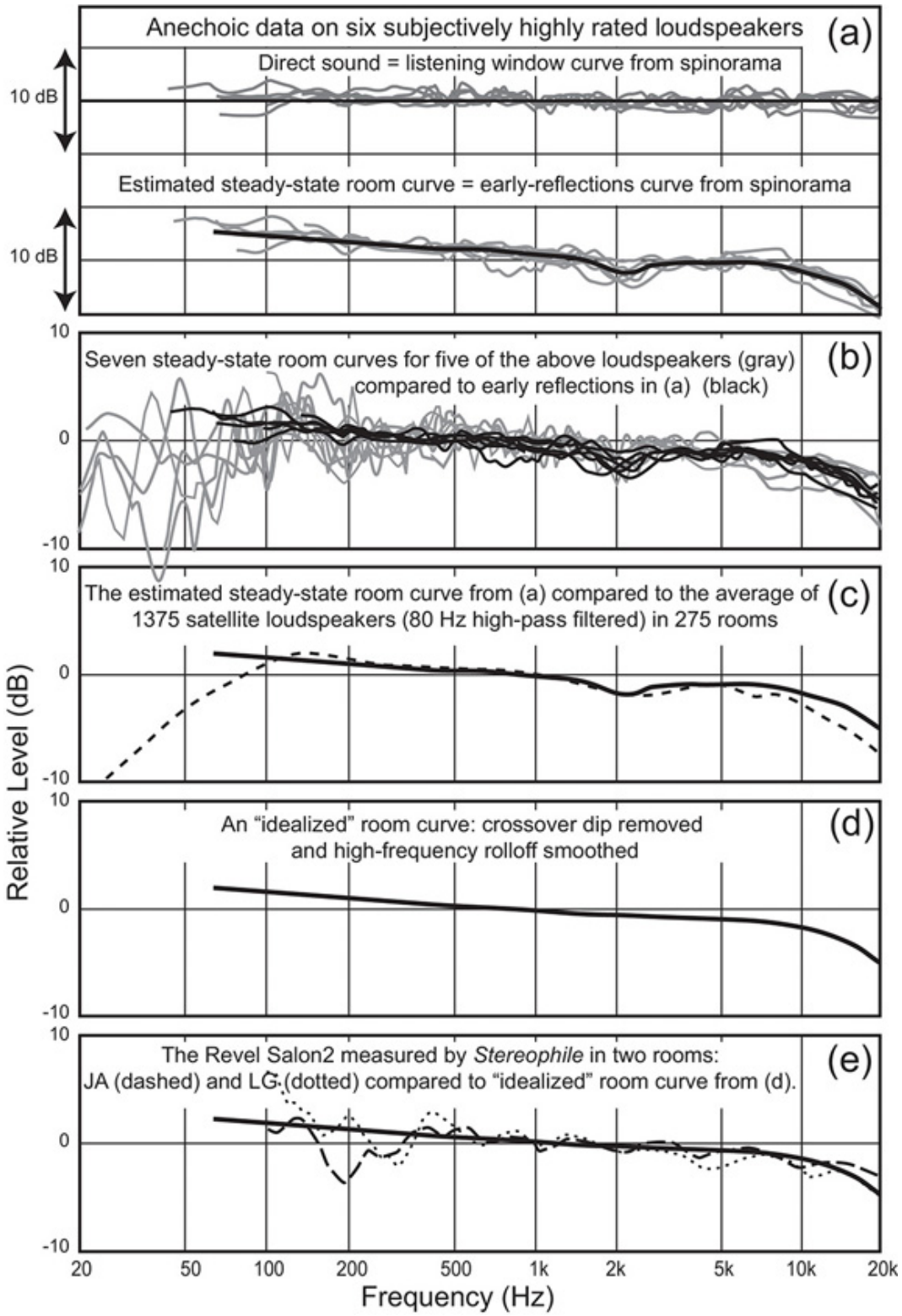


FIGURE 12.16 *Comparisons of anechoic measurements to steady-state room curves. Explanations are in the illustrations. The six loudspeakers in (a) are four Revel loudspeakers: Ultima*

Salon 2, Ultima Studio 2, Performa 3 F208 and Performa3 M106, the JBL Pro M2 in Figure 9.14, and the Infinity Prelude MTS in Figure 2.9. The room curves in (b) are for the Revel Salon2, F208 and M106, Infinity Prelude MTS and JBL Pro M2, measured in five different rooms by four different people using different measuring systems. The survey data in (c) are from Green and Holman, 2010. (d) shows an idealized version of the average room curve prediction shown in (a). The room curves in (e) are from Stereophile magazine, March 2009 (vol. 32, no. 3). Reproduced with permission. ⏎

Figure 12.16(a) shows anechoic data for six loudspeakers that rated highly in double-blind subjective evaluations. The author has drawn a subjectively judged average curve through the anechoic early-reflections data. The dip around 2 kHz is caused by off-axis directivity effects in the midrange-to-tweeter crossovers.

In Figure 12.16(b) the estimated room curves (the early-reflections spinorama curves) in (a) are compared to seven spatially averaged steady-state room curves measured in five different rooms for the Revel Salon2, F208, and M106, the JBL Pro M2 and the Infinity Prelude MTS. All these loudspeakers received high ratings in double-blind listening tests. The measurements include sample variations, room variations and measuring system variations (two of them came from audio forum posts), yet it is evident that the predictions represent a good central tendency for the fluctuating room curves.

The residual fluctuations in the steady-state room curves are acoustical interference effects that the microphone responds to, but human listeners largely or completely ignore. We know from the anechoic data that the loudspeakers do not radiate them. These variations cannot be corrected by "room" equalization. Many room curves are 1/3-octave band smoothed, which would eliminate most of the detailed fluctuations seen here, thereby reducing the temptation to equalize them.

When one sees a *very* smooth high-resolution steady-state room curve after equalization, there is a high possibility that something inappropriate has been done, and the sound quality may have been degraded. Yet this is frequently shown in promotions for room equalization/correction products addressing both consumer and professional customers. This is marketing, not science. The room curve is not the definitive statement of sound quality.

Figure 12.16(c) compares this estimated room curve to data from a very large survey of home-theater satellite loudspeakers (Green and Holman, 2010). The match is excellent: within 1 dB from 100 Hz to about 10 kHz. The crossover dip is evident here as well. These loudspeakers were "as found", without documentation or pedigrees. The roll-off at high frequencies suggests that some of the loudspeakers had rolled-off high frequencies or may have had their tweeters aimed away from the listening/measuring area as in many on/in-wall surround loudspeaker designs.

In Figure 12.16(d) an attempt is made to arrive at a room curve that might result from a "perfect" loudspeaker; one without a crossover dip and one that generated a smoother high-frequency roll-off. The latter differences are minor cosmetics. The result is not a straight line, but a straight line with a slope in the range of −0.4 to −0.5 dB/octave exemplifies the basic information.

In Figure 12.16(e) this curve is compared to room curves measured by John Atkinson of *Stereophile* in his own and another room. The imperfections below about 500 Hz are attributable to the rooms. The small deviations observed in LG's room at high frequencies are puzzling. At these frequencies the direct sound should be the dominant factor. The deviations do not originate in the loudspeaker, so suspicions mount that there might be some acoustical interactions with furniture or walls close to the microphone. Overall, though, the comparison confirms that if one begins with a well-designed loudspeaker the room curves will be close to predictions. Given that anechoic data are definitive, and room curves are subject to variations, the confirmation of excellence is reassuring.

There is no excuse for anyone thinking that the “Harman” curve is a target for anything other than at frequencies below the transition frequency, and even then, one can discuss preferred levels. The problem is that people don’t read. Enough said.

12.3.6 Personal Preferences for Spectral Balance

When loudspeakers are at their final destinations, customers are sometimes not satisfied with what they hear. It is a challenging moment. What happens if the loudspeaker and room issues are settled in the technical sense. Do listeners have individual overriding preferences? If they are given bass and treble controls and the freedom to choose, what would their spectral balance choice be?

Olive et al. (2013) describes an interesting experiment. The authors were interested in determining the relationship between headphone and loudspeaker reproduction. Sean Olive will be discussing headphones in Chapter 15, so I will focus on the results for loudspeaker listening.

The loudspeaker used was highly rated in double-blind evaluations and was commendably flat and smooth on-axis with other well-behaved spinorama curves. Such loudspeakers yield rooms with a gentle downward tilt, as shown in Figure 12.16. For these experiments it was equalized to produce a flat steady-state in-room response, which added a high-frequency boost that, possibly, listeners would recognize and correct.

Three tests were done. In the first two, either the bass or the treble control was adjusted separately with the other one randomly fixed, and a third in which both controls were available, starting from random settings. It was a classic method-of-adjustment experiment. For each program selection, listeners made adjustments to yield the most preferred result.

Figure 12.17 shows the result of evaluations by a mixture of trained and untrained listeners. The four untrained listeners simply choose “more of everything”, adding slightly to the already boosted treble and adding a substantial bass boost. This is the “happy face” spectrum that has a long history in audio. This is very interesting, because in multiple-comparison double-blind subjective evaluations most listeners recognize and reject

resonant colorations, preferring basically neutral loudspeakers. Inexperienced listeners exhibit more variation in their judgments, but their average ratings have been more similar to trained listeners than this result suggests. What was different here?

It was a totally different experiment, not a comparison test. Here listeners were given a very neutral loudspeaker with a smooth treble boost (to achieve a flat room curve) and given the freedom to adjust the bass and treble levels to their taste. The four inexperienced listeners clearly had distinctive preferences. Excessive bass is one of those "forgivable sins", but here it was a choice. In dramatic contrast, the seven trained listeners clearly preferred the loudspeaker in its original form, turning down the treble to deliver a flat direct sound, and doing so with impressive accuracy, Figure 12.17(c). But they too enjoyed a subtle bass boost.

Clearly, customers will use tone controls if they are available and some of their choices can seriously disrupt the circle of confusion and any attempts at program "translation". Many more listeners would need to be tested in order to draw substantive conclusions about specific preference curves, but these data are indicators.

12.3.7 High Frequencies: Air Absorption and Screen Loss

As sound propagates some of the energy is lost at the highest frequencies. The loss is dependent on distance, the prevailing air pressure, temperature and relative humidity. In large concert venues these change as audiences heat the air and exhale humidity. Figure 12.18 shows data that explains what happens in venues and conditions likely to relate to sound reproduction circumstances.

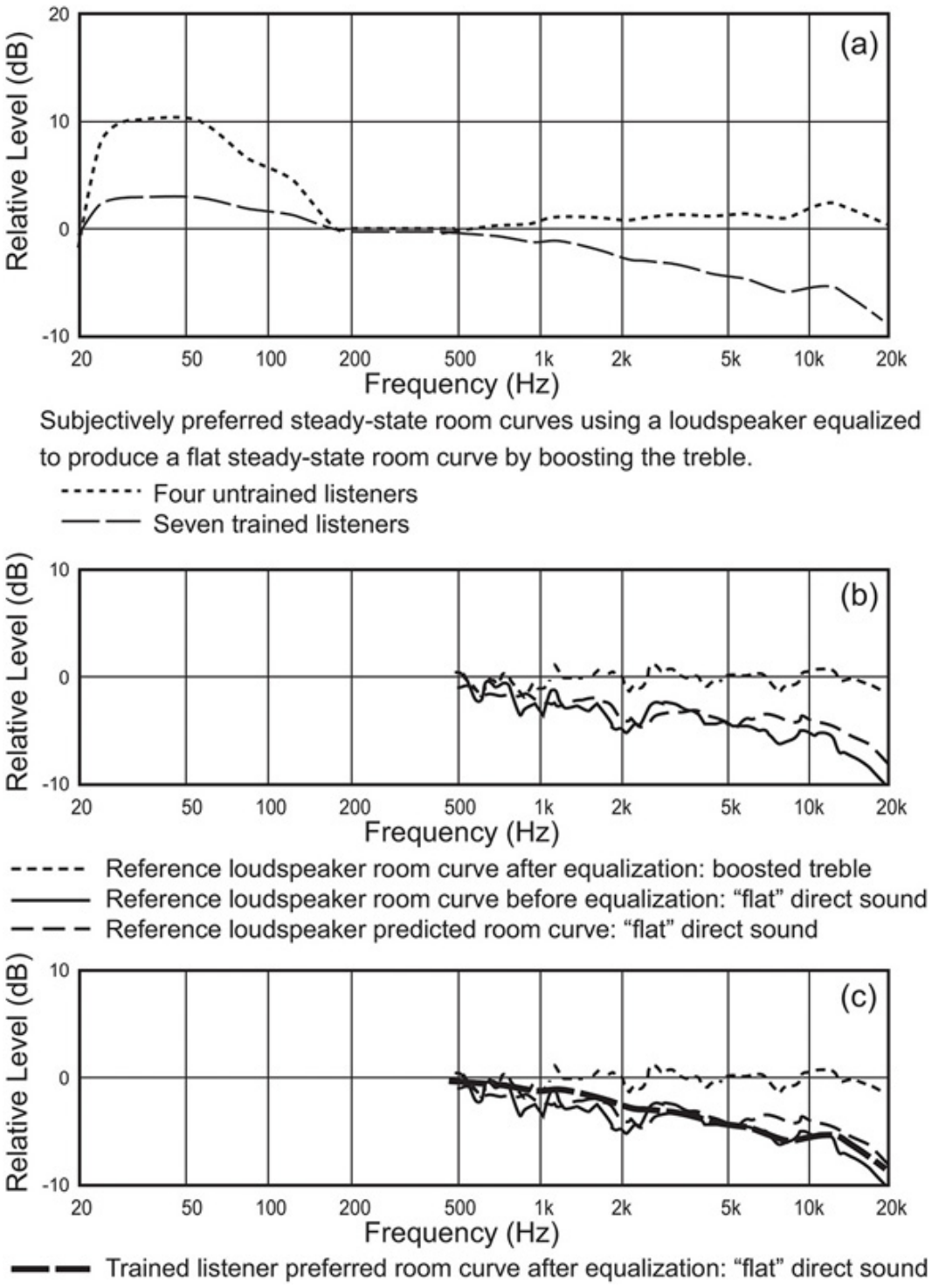


FIGURE 12.17 *(a) Room curves resulting from listeners adjusting bass and treble tone controls to their preferences when listening to a good loudspeaker equalized to produce a flat steady-state room curve (Olive et al., 2013). (b) shows the*

measured room curves for the unequalized and equalized loudspeaker and the superimposed predicted room curve for the unequalized loudspeaker, showing that the prediction from anechoic data is reasonably close. (c) adds the room curve preferred by the trained listeners, from (a). Their tone control adjustments restored the performance of the loudspeaker to its original state, preferring a flat, direct sound. ⏎

This is a phenomenon that we experience daily: sounds at a distance are less loud and sound duller than sounds close up. It is sometimes hypothesized that humans compensate for these effects in the spectra of direct sound components in a sound field. Note that it applies to a single "ray" of sound, which can be the direct sound, or individual reflected sounds, and the amount of attenuation will depend on how far each has traveled. Reflected sounds will have traveled farther and experienced more air attenuation. The effect it has on a steady-state sound field depends on the complex summation of the direct and reflected sounds. In most sound reproduction venues we tend to be in dominantly direct sound fields at frequencies where air absorption is significant, see Figure 2.6(d).

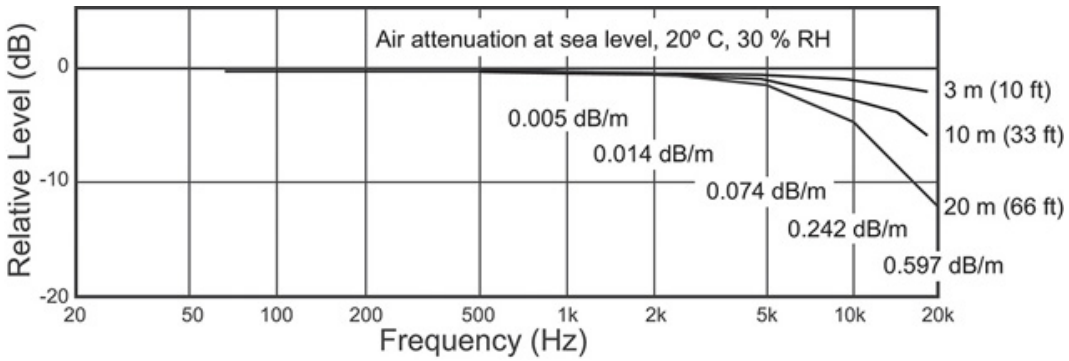


FIGURE 12.18 *High-frequency air losses as a function of distance, and the basic data from which they were calculated. Several internet sources provide calculators for these numbers (e.g. search for "air attenuation of sound").* ⏎

When loudspeakers are placed behind perforated or woven screens there is loss of energy because some of the sound is reflected back towards the loudspeaker. The amount of this reflected energy depends on the thickness and open area of the screen. The losses increase with increasing frequency. There are many options for perforated or woven screens, and they all have some effect at very high frequencies. Remember to include the black light-blocking fabric that may need to be installed to reduce reflections from objects behind the screen. There are trade-offs with video performance that complicate decisions.

Screen manufacturers should provide sound attenuation data, and if not, simply measure with and without the screen in place to evaluate the loss. The measurement should be a time windowed FFT to capture the direct sound, not a room curve. It may be useful to measure at angles off axis, and to vary the loudspeaker-to-screen distance. Sound reflected back at the loudspeaker may need to be absorbed if there are large reflecting surfaces as in a baffle wall.

12.4 ABSORBERS, BASS "TRAPS" AND DIFFUSERS

What we hear from loudspeakers is obviously influenced by the acoustic characteristics at the points of reflection. Figure 12.19 illustrates some popular treatments.

Reflection, Figure 12.19(a), is straightforward—a hard, massive, flat surface generating a specular reflection attenuated only by the inverse-square law, a function of the distance traveled. A hard reflector delivers the strongest reflection to the listener. (b) shows an absorber attenuating the reflected sound by absorbing some of its energy. (c) shows an engineered diffuser sending reduced energy sounds to the listener from many points across its surface, each at a slightly different time. Most of the sound is sent off in other directions, a small amount may be absorbed. (d) shows one of many possible geometric shapes that act as diffusers, here a hemi-cylinder, from which a single, attenuated, reflected ray is sent to the listener, with the

remaining energy dispersed in other directions. Depending on the surface of the hemi-cylinder, some sound may also be absorbed.

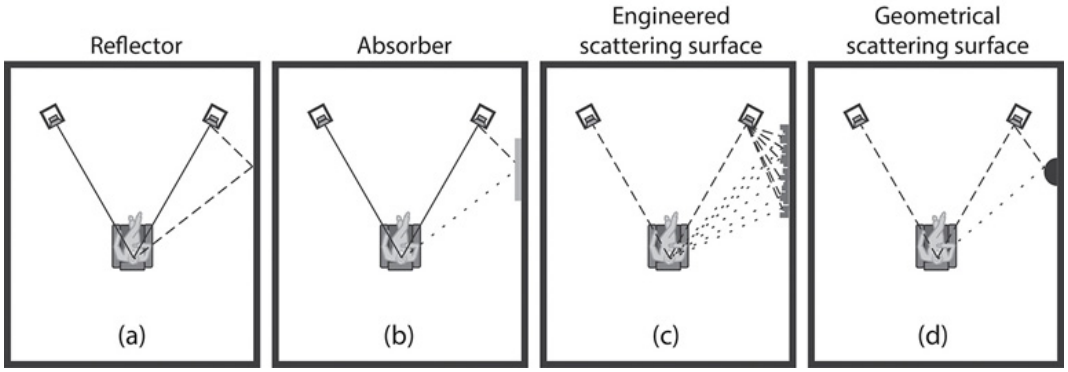


FIGURE 12.19 *shows some popular options for treating the reflection point of a room boundary: (a) reflection, (b) absorption, (c) scattering in time and direction by engineered surfaces (e.g. Schroeder type diffusers) and (d) scattering by geometric shapes (e.g. hemi-cylinders). In (c) and (d) only the scattered components that reach the listener are shown; most of the incident sound energy is scattered in many other directions, depending on the design of the scattering device.*

The picture becomes more complicated when we consider the nature of the reflected and scattered sounds, especially from engineered surfaces. They are likely to be dependent on both frequency and angle of incidence. Precise data are hard to find, especially in a form that is useful when attempting to design a listening room. Not all manufacturers of these devices are equipped to measure them, and specifications tend to be minimal.

12.4.1 Porous/Resistance Absorbers

Section 9.9.3 explained how porous/resistance absorbers like fiberglass, rock wool, and acoustic foam work. Wavelength determines that they are practical to use only at medium-to-high frequencies.

Specifications for the performance of absorbers evolved from Sabine's first work, in large, highly reverberant, diffuse sound fields. The performance of absorbing materials is measured in reverberation chambers, as diffuse a sound field as possible. The reverberation time (RT) of the empty chamber is compared to the RT when a large sample of the material is placed, usually on the floor. What is measured is called the "random-incidence absorption coefficient" because sounds arrive from all possible directions simultaneously. This specification is rigorously standardized. The number is an indication of the relative absorption capabilities of materials, but it is not a percentage. The absorption coefficient often exceeds 1.0. In the small "dead" spaces used for sound reproduction, from homes to cinemas, there is little diffusion in the sound field, so the relevance of the measurement is in question. Still, it is all that exists, so it is used.

The real interest should be in how the material performs when addressing early reflections arriving from specific incident angles. The second loudest sounds to arrive at a listener in typical listening rooms are the early reflections; the first-bounce sounds. We measure the off-axis performance of loudspeakers believing in their importance, but we are missing important information about the performance of absorbers at the relevant angles.

Beginning with standard specifications, Figure 12.20 shows the random-incidence absorption coefficient for 1-, 2-, 3- and 4-inch thicknesses of fiberglass board.

This figure shows that this material has increasing amounts of low-frequency absorption as thickness increases, and the numbers imply "perfect" absorption above about 1 kHz. The absorption coefficients regularly exceed 1.0, indicating that the specification is not a percentage absorption. There are important messages in these data:

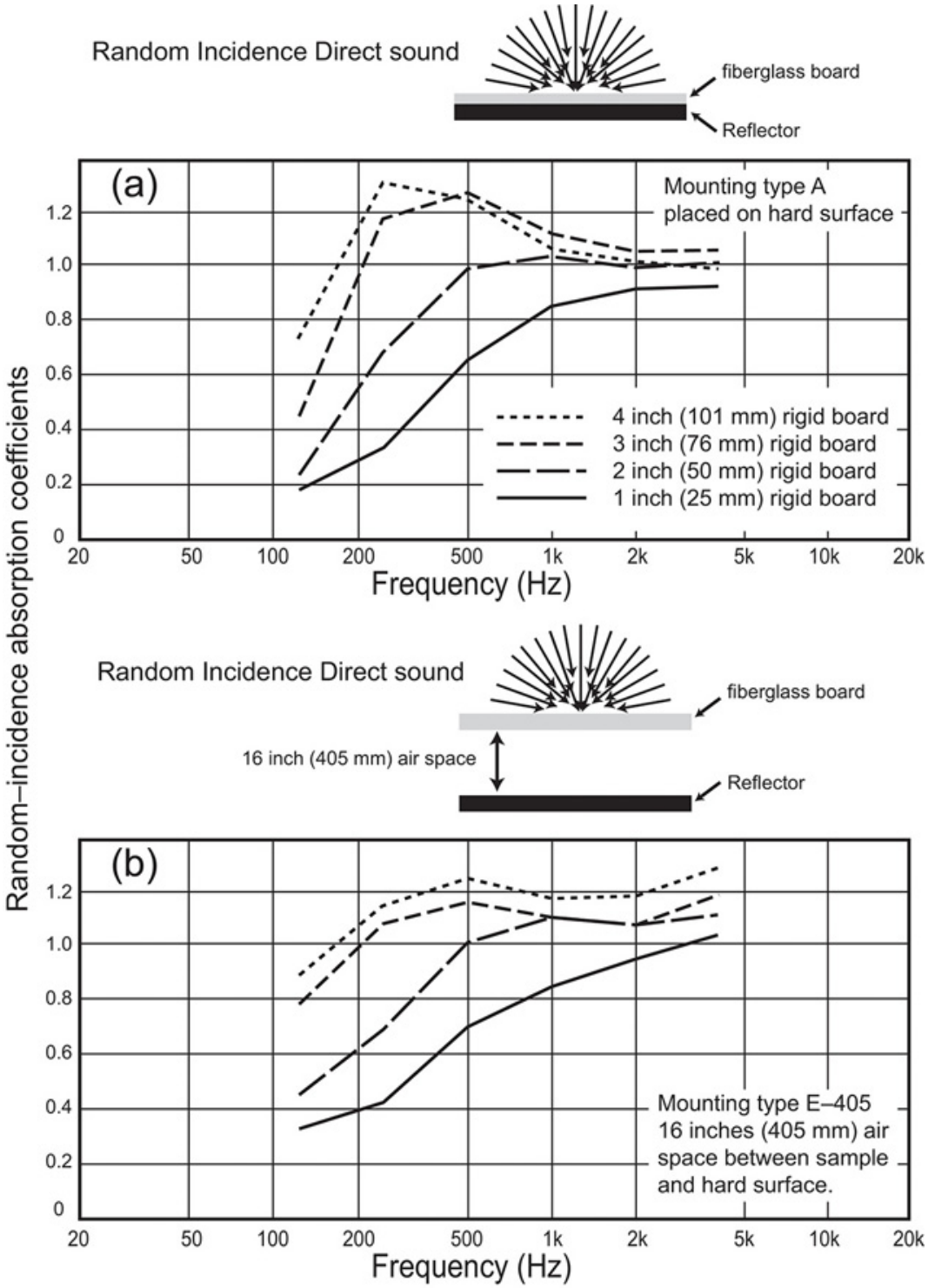


FIGURE 12.20 *(a) shows the published random-incidence absorption coefficients for various thicknesses of fiberglass mounted in a common fashion, directly on top of a reflecting*

surface. (b) shows the effect of adding an air space behind the fiberglass board. Data from www.owenscorning.com. ⏎

1. As thickness increases, low-frequency absorption increases.
2. If we wish to absorb sound down to the transition frequency, absorbers need to be 3 to 4 inches (76–101 mm) thick.
3. Effective thickness can be increased by incorporating an air space behind the fiberglass panel. Filling the space with low-density fibrous material is beneficial.

Not stated is the important fact that the density of the fiberglass board has negligible effect on the absorption coefficient, but a large effect on the mechanical strength. Anechoic chambers use the lowest density that can hold the shape of the wedges, often about 3 lb/ft^3. My understanding is that highest density boards were developed for insulating flat roofs, where workers walked on them while installing tar and gravel waterproofing. Now foam plastics tend to be used. In room acoustics stiff boards make it convenient to wrap fabric around them, and with the strength to resist dents. Otherwise, even low-density household insulation is useful. The performance of installations shown in Figure 12.20(b) can be improved by using a thin surface board and filling the void with inexpensive fluff. Obviously, any space is a low-cost improvement, not just the standardized distance shown.

Only recently have researchers addressed the issue of directional absorption, the performance that matters when we place absorbers to reduce the energy in specific early reflections, like the side wall reflections. The angle of incidence of these reflections is often in the range 45° to 60°, as can be seen in Figure 2.10. Across-the-room direct sound from surround loudspeakers may approach the absorber at near normal incidence, perpendicular to the surface. Figure 12.21 shows what might be happening at those angles.

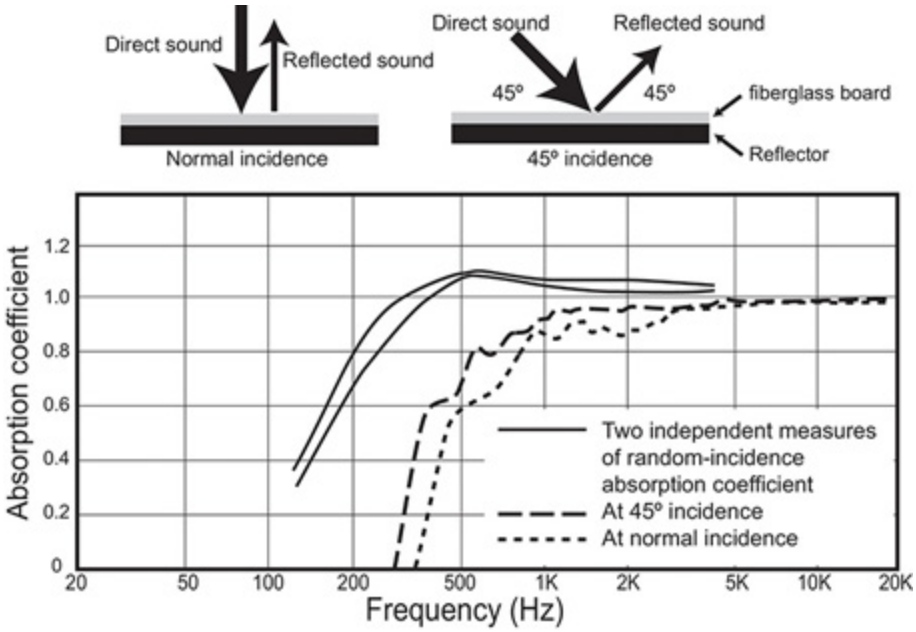


FIGURE 12.21 *Results from absorption coefficient measurements at normal (0°) and at 45°, compared to random incidence measurements on the same material. Redrawn from Rathsam and D'Antonio (2006).* ⏎

Figure 12.21 is persuasive evidence that the angle of incidence matters greatly to the acoustical absorption of fiberglass panels. The low-bass performance is seriously degraded, and the absorption coefficients do not exceed 1.0. The absorption at normal incidence is lower than at 45° because the sound travels a shorter path to the reflecting surface. Lower incident angles experience greater absorption, explaining why random incidence measurements are highest because they include all angles up to grazing incidence, where the absorbing path length is maximum.

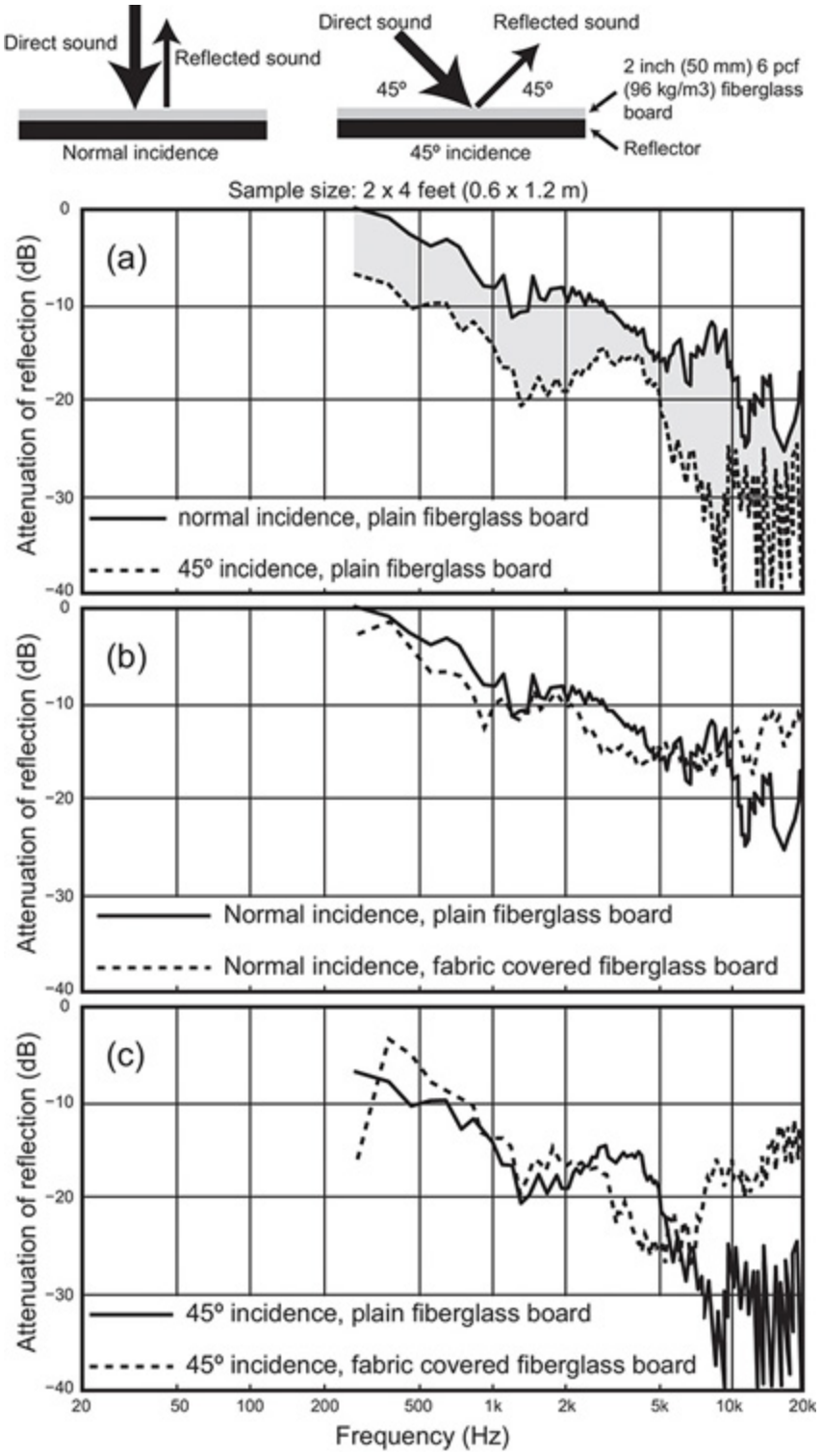


FIGURE 12.22 *(a) shows the frequency-dependent attenuation of a single reflection from 2-inch (50 mm), 6 lb/ft^3 (96 kg/m^3) fiberglass board. Reflections are shown for normal and 45° incident angles. The difference area is shaded. (b) shows the*

attenuation at normal incidence for the plain fiberglass panel and with the panel covered with a popular acoustical fabric, Guilford of Maine FR701. (c) shows the comparable data for a 45° incident angle. Data courtesy of Peter D'Antonio. ⏎

This is important information, but more directly related to our interests is the information in Figure 12.22 showing frequency responses of the reflected sounds at specific incident angles.

These measurements again show increased absorption at 45° incidence, which forces that sound to travel through more fiberglass. But the most interesting observation is that the effect an absorber has on reflected sound spectrum is not predicted by the random incidence absorption coefficient. The effect, in reality, is to roll off the high frequencies, not to eliminate the reflection. The roll-off is quite uneven at 45°, making the early-reflected sound field different from what might be anticipated from off-axis anechoic loudspeaker data.

The fabric cover has only a mild effect over most of the frequency range but becomes reflective at the highest frequencies. The published random-incidence absorption coefficient for acoustical panels will not show this effect because the measurements typically stop at the 4 kHz octave band, which has a high-frequency limit of 6 kHz, just avoiding the reflected sound shown here.

Providing a different perspective on the same phenomenon, Cuharero et al. (2020) examined the reflective/absorbing properties of a variety of materials. Figure 12.23 shows my rendering of their findings for sound reflected from 2-inch (50 mm), 3.2 lb/ft^3 (52 kg/m^3) fiberglass.

Figures 12.21 and 12.22 show that the sound reflected from a flat fiberglass panel is both angle and frequency dependent. Figure 12.23 confirms this from a different perspective. Incidentally, these two measurements happened to use different densities of fiberglass, but as mentioned earlier, absorption changes little with differences in density.

These reflected-sound measurements emphasize the need for very thick absorbers if the intention is to remove the reflection, not just change its

timbre. Thick fiberglass panels are expensive, so adding a fluff-filled space behind may be worth considering. The assembly will need to be enclosed within a frame enclosing the edges for maximum effect. This will move the curves downward in frequency, thereby absorbing sound over more of the frequency range. The widely used 1-inch (25 mm) fiberglass or acoustic foam board, or 2-inch sculptured foam (half of the volume is air) perform similarly and are definitely not advised—except as acoustically translucent facings for an absorber employing a fibrous-filled space behind.

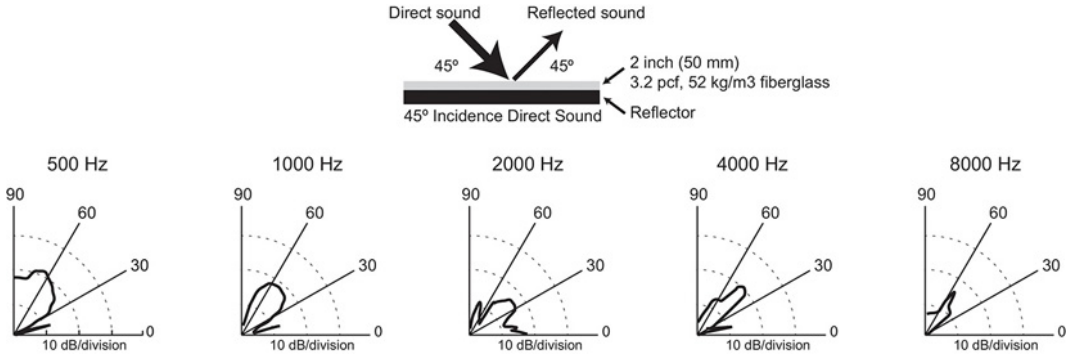


FIGURE 12.23 *Directional reflection patterns from 2-inch (50 mm), 3.2 lb/ft^3 (52 kg/m^3) fiberglass at 0.5, 1, 2, 4 and 8 kHz. Information extracted and redrawn from figure 6 in Cuharero et al. (2020).* ⏎

In custom listening rooms it is common to cover large wall surfaces with stretched fabric to cover what might be otherwise an unattractive collection of diffusers, absorbers and loudspeakers. This is more than a visual alteration to the room because fabric has flow resistance; if you cannot easily see daylight through a fabric, or blow through it, it will not be acoustically transparent. Such fabrics, with nothing more than empty space behind them, become absorbers, and experience shows that many listening rooms end up being unpleasantly dead because of it. I realize that such an effect is initially impressive to new custom home theater owners—it is a choice. See Figure 12.26 for an example of an ornamental covering creating a frequency-selective absorber.

12.4.2 Diaphragmatic/Membrane Absorbers: Bass "Traps"

Section 9.9.3 provides the background to this type of low-frequency absorber. The common name, bass trap, is suggestive of performance that is rarely realized. It seems to have originated in the recording industry. Whatever the real history, the reality is that attenuating/absorbing bass energy in rooms is a challenge, and we wish we could simply "trap" it.

Acoustical fiber and foam absorbers need to be impractically thick to be effective at bass frequencies, so surface areas that move in response to sound, and dissipate energy while moving, are used. They only work when located in high pressure regions of the problematic standing waves, so the first step is to identify those frequencies. This is a possible procedure:

1. Identify the problematic resonance frequencies. This requires measurements, and the results will depend on the location(s) of the active loudspeakers and the location(s) of the important listeners.
2. Locate the high-pressure regions on the room boundaries where absorbers can be placed to be maximally effective for the identified standing waves. Corners are good locations for addressing multiple resonant modes.
3. Find manufactured absorbers, or design, build and test your own absorbers that are effective at the problem frequencies. Be cautious about manufacturer's claimed performance—it is expensive and difficult to perform these measurements with any precision.
4. Hire an interior designer to create a pleasing scheme that incorporates these large, inherently unattractive devices. Recording control rooms and custom home theaters are dedicated to acoustical/visual performance, and normal concepts of room décor do not apply.

Very early in my career, when I was setting up the first NRCC listening room in the late 1960s I needed tuned low-frequency absorbers to attenuate strong standing waves in the concrete- and plaster-surfaced laboratory

space. I could find no commercially available products, so, using what little information I could find as guidance, I built some prototypes and tested them. Lacking an accelerometer, I used a sturdy inexpensive phono cartridge as a vibration sensor with negligible moving mass. Sweeping pure tones through a nearby loudspeaker excited the panel and the trial-and-error process began. Progress was monitored by measuring RT and room curves. It worked remarkably well, but in that room, six 4 ft × 8 ft × 6 inches and 8 inches (122 × 244 × 15 cm) membrane absorbers were needed to get RT into the range of normal domestic rooms. It was a time-consuming but excellent way to learn about low-frequency room acoustics. The room ended up functional but unattractive.

It is likely that this experience encouraged me, years later, when I had a bass resonance problem in my home listening room, to seek other solutions. This led to my first visually unobtrusive multiple-subwoofer manipulation of room modes, and Todd Welti's subsequent work to be described in Chapter 14.

It is fortunate that a single layer of drywall is a useful membrane absorber, easing the problem in many typical homes. Obviously, any amount of low-frequency absorption is beneficial, even if the intention is to use multiple subwoofers. Some enthusiasts have abandoned all concerns about interior décor and covered most surfaces of their listening rooms with large rectangular low-frequency absorbers. I have seen photographs. I hope they found happiness and tight bass. Active bass absorbers are an effective and visually inconspicuous option.

However, with a well set up multiple subwoofer installation, bass traps may not be needed. It is a trade-off of cost, convenience, appearance and function. Multiple subs; for most rooms they don't have to be large, and there are clever designs that are not visually offensive, and sometimes can be hidden. Will be discussed in Chapter 14, absorption of any kind reduces system efficiency; multiple subwoofer arrays usually increase system efficiency. There are choices.

12.4.3 Geometric-Shape Diffusers

The original source of scattered sound in concert venues was irregular architectural shapes, opera boxes, and statuary. Modern halls use engineered irregular and curved surfaces. Diffusers must manipulate sound waves of significant size and must therefore be of comparable size. One of the early popular shapes was the hemi-cylinder or, in a slightly flattened form, the polycylindrical absorber/diffuser. Large ones were covered with thin membranes so that they also served as absorbers in the upper bass region.

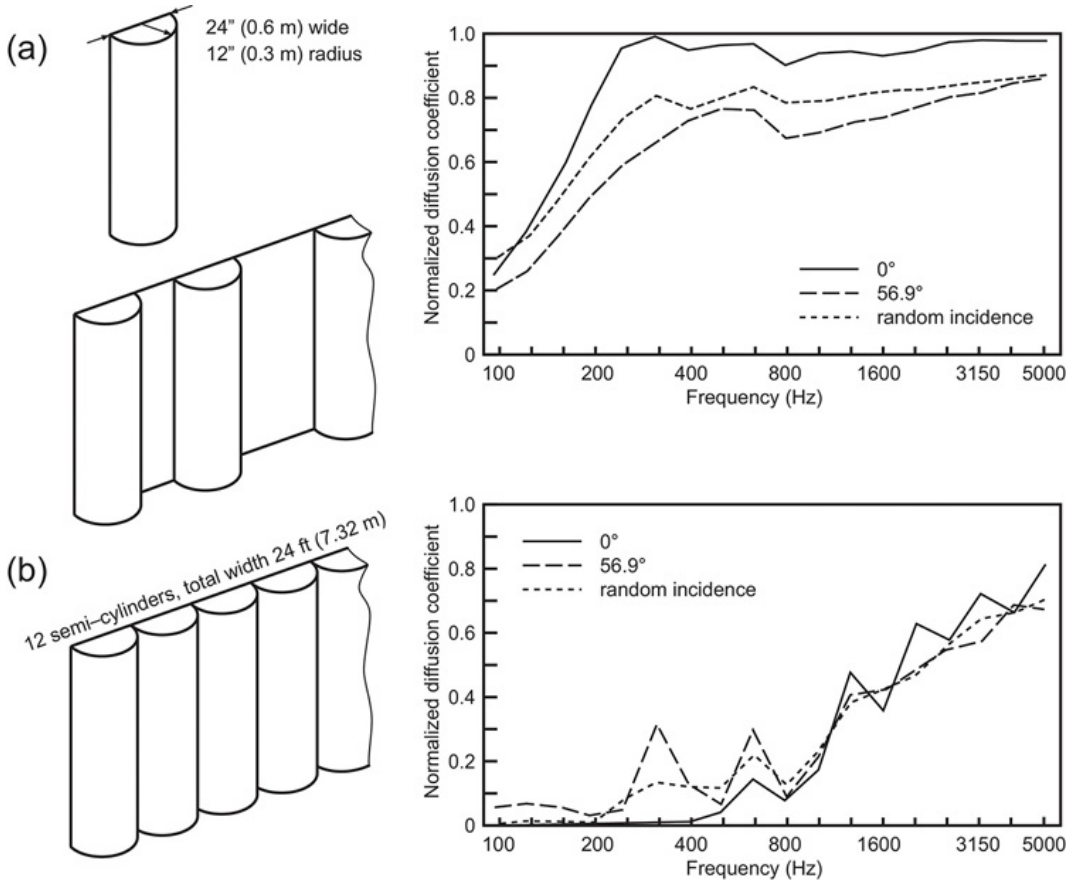


FIGURE 12.24 *(a) The normalized diffusion coefficient for a single hemi-cylinder and randomly spaced units. (b) Adjacent hemi-cylinders covering a surface. Data courtesy Peter D'Antonio.* ⏎

Figure 12.24 shows the normalized diffusion coefficient for a single, and then for multiple hemi-cylinders with a 1 ft (0.3 m) diameter. The devices work well in isolation and when randomly spaced, showing Normalized

Diffusion Coefficients between 0.7 and 1.0 above about 300 Hz, to a usefully low frequency. However, when placed contiguously, touching each other, the low-frequency diffusion is lost—the geometric regularity is also visible in the cyclical pattern in the curve. As Gilford (1959) noted, other multifaceted shapes also work well, but sizing and placement are important so that they influence the intended reflections—see Figure 12.25.

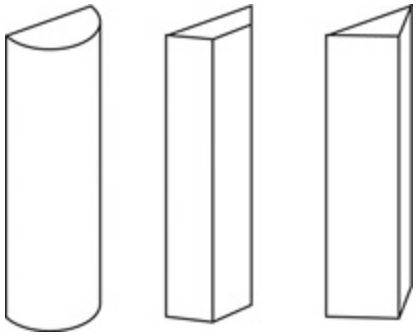


FIGURE 12.25 *Many artistic variations are possible for diffusing columns, but acoustically the hemi-cylinder is best.*

For home listening spaces columns are an interesting, and in some custom rooms, a decorative option (Figures 8.14 and 12.29). Surround loudspeakers can be hidden in them. Of all possible shapes, the hemi-cylinder is most desirable for its uniform scattering of incoming sounds from many directions. Other curved shapes, including flattened semi-cylinders, work, but simply not as well (Cox and D'Antonio, 2009). To cover the frequency range down to the transition frequency, it seems that a depth of about 1 ft (0.3 m) is required.

12.4.4 Engineered-Surface Diffusers

These devices now appear in countless versions, many of which are clearly designed for visual appeal rather than acoustical performance. It can always be claimed that irregular surfaces scatter sound, but the reality is that some designs do it much better than others. If one cares about the performance, it will be difficult to find meaningful specifications because they take time

and money to measure. Visual appeal is powerful, and frankly, almost any irregular surface is sufficient to address flutter echoes.

A digression: the correct spelling is *diffuser*. "Diffusor" is the spelling chosen by RPG to differentiate their products. Not everyone got the memo.

Engineered surfaces, with deeply sculptured faces calculated using number theory and other concepts are versatile devices, able to disperse sound in a single plane or multiple planes. Figure 12.26 illustrates the performance of some commercially available diffusers. These devices also absorb some sound, and this is shown in the lower graph. If the installer decides to cover the diffusers with fabric to make them more compatible with the visual décor, the absorption increases substantially and becomes strongly frequency dependent—a bad thing.

These devices are indisputably useful for adding complexity to sound fields in performing arts venues, but in small sound reproducing spaces they serve a different purpose—to attenuate early reflections, not to create a more diffuse sound field. These surfaces are designed so that incoming sounds are scattered in many directions from all parts of the device, so only a portion of the incident sound energy is delivered to the listener. The sound sent to the rest of the room is of little interest because we do not want, and cannot have, a diffuse sound field in a typical listening room.

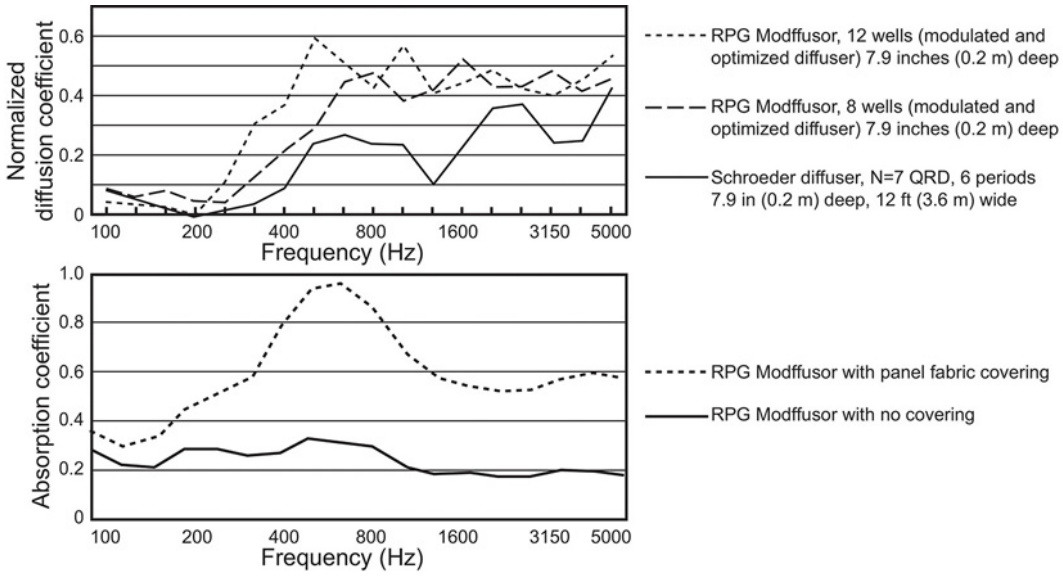


FIGURE 12.26 *A comparison of the normalized diffusion coefficient for a classic Schroeder QRD diffuser with an evolved form, the RPG Modffusor, that has been modulated and optimized for improved performance. It can be seen that the Modffusor improves when used in larger forms. Also shown is the absorption coefficient for the Modffusor with and without a panel fabric covering. Data courtesy Peter D'Antonio and www.rpgacoustic.com.* ⏎

The concepts of diffusion coefficient and scattering coefficients may be foreign to some readers, but they are relevant to defining the performance of these devices. Cox et al. (2006) provides an in-depth explanation.

Figure 12.27 shows that incoming sound from a loudspeaker at 30° incidence is redirected almost uniformly over a 180° angular range. The flat surface reflects the bulk of the energy towards the listener at the symmetrical angle. In this demonstration the effective attenuation of the reflected sound is an impressive 20 dB. In normal rectangular listening room setups the first side wall reflections tend to be in the angular range 45° to 60°, so more specific information is needed about angle dependency, and some information about the spectrum of the diffused sound that is sent in the direction of the listener would also be welcomed. Any such data would obviously apply only to specific devices and ideally should be made available by the manufacturer.

These surfaces are really distributed sound sources, delivering many "mini-reflections" to the listening position while scattering the rest of the sound to other parts of the room. This has been thought of as an advantage, as it leads to an improved appearance of ETCs (energy-time curves) is often used as proof.

Some research indicates that the perceptual sum of several mini-reflections is equivalent to a larger single reflection. This goes back to Cremer and Müller (1982) (see Figure 6.14). Angus (1997, 1999) provides confirmation and added to by Robinson et al. (2013). This means that the time-domain data does not convey reliable information about audibility of

reflections—replacing a high spike with a collection of smaller ones may sound more similar than is implied by the visible data. It continues to be a topic for more research in the present context. To function down to the transition frequency these devices need to be about 8 inches (200 mm) thick, as seen in Figure 12.26.

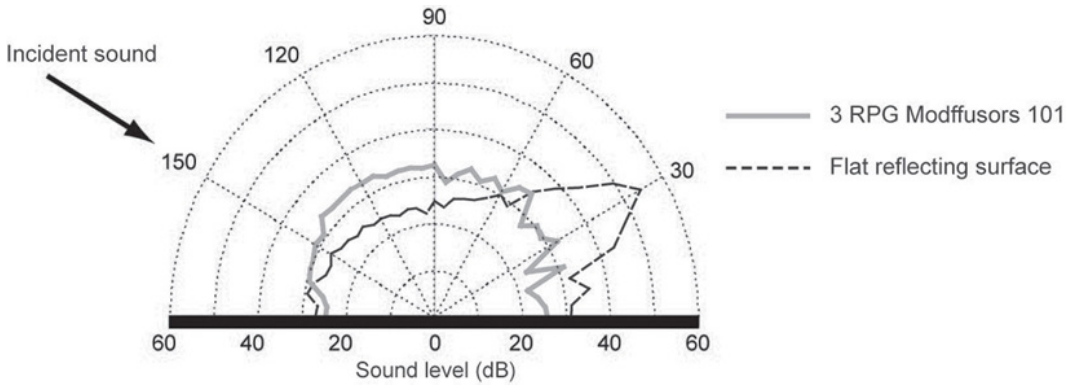


FIGURE 12.27 *The distribution of sound from one design of an engineered diffuser compared to a flat reflecting surface.*

In the era during which it was fashionable for the loudspeaker end of a control room to be acoustically "dead", and the opposite end to be "live", I toured an elaborate recording complex that was nearing completion. I had not experienced one of these new rooms and was given a demonstration. I sat at the center of the console and listened; something was wrong, the center image was strangely unclear. Looking around I saw that the entire back wall was covered with beautiful and expensive wooden diffusers, vertically arranged to spread the sound horizontally. I asked for monophonic pink noise to be played, and from the stereo seat there was simply no center image—the sound spread across much of the space between the loudspeakers. Others heard it too; it was not subtle. There was concern, as this was obviously not good. The movers were still on the premises, so we found some moving blankets and hung two of them in front of the diffusers. The center image snapped into place where it belonged. The problem: too much uncorrelated sound arriving from behind the listener, scrambling the information in the direct sound from the

loudspeakers. Lesson: there can be too much of a good thing. The ETC probably looked impressive, though. Somebody may have measured but they did not listen.

This is why, in general, unless they are far behind the mix position, diffusers are best located on the side walls or ceiling (sending the sound energy towards the sides), with absorbers on the front and rear walls. Section 12.5.3 describes a room that warrants some psychoacoustic analysis.

12.5 LOUDSPEAKERS AND ROOMS DETERMINE "TRANSLATION"

The confused situation of recording control rooms was discussed in Section 3.5. There is no agreed standard, and yet these are the places where the recorded arts are created, where imagination, inspiration, musical composition and performing talents are combined and manipulated by other artists—recording engineers—into master recordings that will be reproduced countless times in countless locations. The transient experiences in control rooms influence what become permanent features of the art. Recording control rooms are also showplaces, part of marketing services to clients. Proof of success is a list of popular musicians who have used the facility, as if it is the facility, not the music, talented musicians, or the recording engineer, that caused the success. Successful art finds itself attached to physical things, rooms or loudspeakers, or to individuals involved with its creation. In reality, it is a collaborative achievement. If the art succeeded in the marketplace, it must have "translated" to the masses. What was responsible?

The main theme of this book has been sound quality, and it has been conclusively shown that the loudspeaker is the prime determinant of sound quality, just as in live performances the voices and musical instruments are the prime determinants of our perceptions. Rooms matter, but unless there are obvious acoustical abnormalities, they become backgrounds perceptually integrated into the performances.

Most recordings are abstract creations, incorporating hints of live performances, but creativity prevails, as it should. The chosen format is an important factor. Stereo, the default format for most programs, is directionally and spatially deprived. It adds insult by delivering the featured musical artists on the soundstage using pan-potted double-mono with its inherent interaural crosstalk distortions (see Section 4.3). A phantom center image should not sound different from a real center loudspeaker if there wasn't a problem (King et al., 2023). Professionals and consumers have both struggled mightily to overcome the limitations of stereo, using signal processing, innovative loudspeaker configurations, room acoustical tricks and adding upmixed channels.

Multichannel audio is a great improvement. Persuasive envelopment can now be enjoyed, and movies have off-screen audio opportunities. Still, only a tiny fraction of listeners are equipped to enjoy it through loudspeakers. Immersive systems add excitement to blockbuster movies, and can add spatial subtleties to music, but the loudspeaker audience for these formats is even smaller. It comes down to cost, complexity, available space, and interest. Headphone binaural or crosstalk-cancelled two-loudspeaker versions may yet prove to be rewarding and practical delivery formats for the masses (see Section 3.3).

This is a complicated and not very flattering situation for a mature audio industry that has had decades to get its house in order and has not succeeded. Without a scientific foundation it could not, and unless the inclination to believe in scientific facts improves, it never will.

The concluding remarks of Fazenda et al. (2015), quoted in Section 3.5, indicate that choosing a recording control room design is a personal choice, and there are several distinctive, incompatible, options. We know that personal choice has a long (unimpressive) history in the choice of monitor loudspeakers (see Sections 3.5.1 and 3.5.2). Now there are facts that can add value to the process.

12.5.1 A Perspective on the "Translation" of Recordings

Fundamental to the entire entertainment industry is the desire to have the "experience", whatever it is, "translate" to the audience. Before that, there is a need for translation within the industry, because recordings can evolve over time, involving several facilities and persons. The translation internal to the industry requires similarity of sound, but that sound can be arbitrary. The translation of the recordings to the outside world cannot be arbitrary.

The concept had its most practical application in the movie industry, described in Section 3.6 as a "closed loop", within which industry standards attempted to ensure that cinema attendees heard and saw what movie mixers created in their cinema-like dubbing stages. But now, movies are played everywhere: homes, planes, tablets and smartphones, often viewed on tiny displays and heard through headphones. They are compromises forced upon us by circumstances, but a portion of the art does translate in a distorted form; it is still entertainment.

In contrast, the music industry is an "open loop". There are no standards for any part of it. Monitor loudspeakers over the years have ranged from rubbish to excellence, and recording control rooms have taken some forms that no reasonable person would want to replicate in their homes. Consumers listen through anything, large or small, fixed or portable, loudspeakers or headphones. And yet there is concern about "translation". Music has always survived being delivered through seriously flawed loudspeakers, in sometimes noisy backgrounds. We grasp hints and imagine the rest. Something "translates", though much is missing; it is still entertainment. Manufacturers of sound reproducers routinely "voice" their products in attempts to best "translate" recordings that their customers may be listening to.

Obviously, accurate translation is not happening. Do we abandon all hope, or do we dedicate our efforts to the audience that is able to appreciate the art that is created? The latter approach must be followed to motivate those who create the art. Fortunately, this audience exists, and it is getting larger. These days even small inexpensive devices can exhibit essentially neutral, albeit band- and loudness-limited performances. Inexpensive well-

chosen headphones can be superb. Moderate cost, well-chosen home audio systems can be very satisfying.

My colleagues and I are in the science culture, but we appreciate the arts, which is why we look for explanations for what we hear, and how it may be possible to make listening experiences better, and more consistently good. We understand a lot about listener preferences, and how to set performance objectives for loudspeakers and headphones that generate high ratings in double-blind listening tests. It is important to ask, what have we been listening to?

> The 60 years of successful research reported in this book has used commercial recordings as source material. The remarkable thing is that it all has "translated" well enough to allow listeners in double-blind evaluations to reliably identify the most technically neutral, most perceptually transparent, loudspeakers that exist, rejecting or downgrading lesser products. How can this be?

The answer is that in well-designed subjective evaluations listeners reject loudspeakers with audible resonances because they modify the timbre of the voices and instruments in music. Good microphones, electronic processing, and most spectral balance adjustments in the recording process do not add the narrow-band resonances that listeners object to. They do not add the frequency-dependent directivity flaws in loudspeakers that listeners object to. They do not restrict low-frequency extension which listeners would object to. They don't add the annoying bass booms resulting from listening room resonances. In short, most recording engineers been doing what is necessary for the essential *sound quality* in their products to translate very well indeed.

Unfortunately, they have not always been hearing the best of their own products because of flawed monitor loudspeakers (see Section 3.5.1) and atypical listening spaces. The topic is alive in both professional and consumer audio forums; there is some cross-fertilization. This is as it

should be because a good loudspeaker is a good loudspeaker, as long as it plays loud enough and long enough for the purpose.

12.5.2 Close Listening Is Useful, But …

The arrival of close-listening, near-field monitoring was inevitable. The little Auratone 5C of the 1960s was a translation reference for a common form of "bad" sound of the period, which got extended in the Yamaha NS-10M (Section 3.5.2). But bad sound comes in countless other forms and, fortunately, has been diminishing. The evolution of the small loudspeaker on the meter-bridge to the present popularity of close listening is logical and sensible. It puts the engineer in a dominant direct sound field, away from the complexity and idiosyncrasies of control rooms. It also might lessen some side effects of hearing loss (Chapter 16). With the trend away from large reflective consoles it is possible to elevate small loudspeakers to ear level, minimizing nearby reflecting surfaces, and deliver a relatively pure direct sound to the listening position. It fits naturally into the current trend to home studios.

There are some excellent neutral small loudspeakers in circulation. In close listening the control room itself is substantially ignored, and translation to other control rooms is much more predictable. Just take the little loudspeakers with you. If the near-field monitors are neutral, the translation of sound quality to good home systems and good headphones is also likely to be successful.

The disadvantage of close/near-field listening is that all the flaws in stereo are maximally audible. There are no local room reflections to moderate the colorations created by acoustical crosstalk and hard-panned, mono, lateral images (see Sections 4.3.1 and 4.3.2). Human adaptation is amazing.

Mid-field, behind the console, locations are good representations of what many audiophiles, among the most severe critics, hear. Some room sound will be included, so more care in setup is required, but the translation then includes aspects of the listening space, which many customers find attractive. The perception of large spaces in recordings will likely dominate,

so it is not necessary to "eliminate" the listening room (Hughes et al., 2016).

12.5.3 Translation Viewed From the Reproduction End

When I joined Harman International in 1991 as Corporate Vice President of Acoustical Engineering, it was after 27 years as a research scientist in the National Research Council of Canada. There are huge differences between those jobs. In one situation I was free to investigate, learn and report results of experiments. I was the judge. In the other, I would be judged on the basis of how Harman's products were perceived in the marketplace, and we had many products in many marketplaces. Our products had to "translate" programs from many sources to listeners of many kinds. Manufacturers of playback devices don't expect recordings to be customized for them, and none are. They do their own surveys of "typical" program sources and attempt to deliver attractive sound quality in their products.

In Harman I created a corporate research group that conducted controlled, double-blind subjective evaluations to confirm that there was subjective acceptance of many of the products being sold. A few very experienced loudspeaker engineers resented having their judgment questioned, even though measurements indicated that their designs were not consistent. An elaborate blind vs. sighted test demonstrated just how compelling non-auditory factors were in the formation of opinions. Experience and confidence in one's opinions did not matter, everyone was susceptible to bias (see Section 10.1). Double-blind subjective evaluations became part of the design process. In fact, our process of verification was sufficient to terminate misleading loudspeaker reviews in *Consumer Reports* magazine (see Sections 11.0 and 11.1).

> In effect, we had confirmed that commercial recordings used in the evaluations translated the essentials of sound quality very well. The artistic and spatial contents remained to be judged, and they are different dimensions of "translation".

Being realistic, manufacturers give customers some freedom to adjust for personal taste, acoustical and background noise characteristics of the listening circumstances, and, of course, variations in programs. Rental cars often provide evidence of the extreme use of tone controls, home stereo systems are often distinctively "calibrated" on site and most have tone controls, some streaming services allow tone control adjustments, even elaborate parametric equalization filters. Headphones now have smart phone apps to customize listening experiences. Next-generation entertainment schemes all provide for customer control of many parameters. Every adjustment of these kinds is another step removed from a reliable indicator of what customers are hearing—pertaining only to spectrum.

12.5.4 Translation Recommendations

From the preceding discussion it is clear that recording engineers cannot anticipate what customers are likely to hear. There are too many variables. So, how does one estimate what will translate? There are, it seems, three fundamental components: sound quality, directional/spatial effects and the artistic content. Only the first two can be physically controlled, and evidence is that they are comparably important to our overall satisfaction. Here are suggestions for loudspeaker listening. See Chapter 15 for guidance on headphones.

1. Select timbrally neutral monitors, because these are the loudspeakers listeners prefer in double-blind tests when listening to a broad base of commercial recordings. This is a demonstrable trend in professional monitor and consumer loudspeakers, as well as soundbars, small bluetooth loudspeakers and headphones. This book describes how to find them, and proves the value of good technical data. Comprehensive anechoic data, competently interpreted, are reliable guidance. Listeners are easily biased in sighted tests, see Toole and Olive (1994) and Section 10.1. Perceptually, the direct sound is the dominant factor in sound quality evaluations and localization. Steady-state room curves

are not reliable information *above* the transition/Schroeder frequency, in spite of tradition and marketing claims. They are results, not targets for equalization.

2. Correct the bass response to attenuate prominent room modes and set a reasonable balance. Bass is about 30% of our overall impression of sound quality. Steady-state room curves are reliable data *below* the transition/Schroeder frequency. Chapter 14 describes how to deliver good bass to one or multiple listeners.
3. Minimize the effects of *objectionable* room reflections. ITU and EBU standards recommend attenuating *all* early reflections (including the floor?) occurring within 15 ms by 10 dB at frequencies between 1 kHz and 5 kHz. This unbalances the spectrum of the reflected sound field by attenuating high frequencies, an unwise action. Reflections should be absorbed or attenuated uniformly down to the transition/Schroeder frequency. Achieving this in a control room can be expensive, and it may not be done in home studios. Close listening can achieve a comparable result at much reduced cost. Just be aware that in any direct sound field listening the flaws in stereo will be maximally audible for the listener in the sweet spot (see Section 4.3). Avoid compensating for the 2 kHz dip in the center phantom image using equalization. The problem is comb filtering—a non-minimum-phase problem—and EQ is not a solution. It adds a resonance to the playback that will be audible everywhere except the sweet spot. It will also be heard as a spectral error in headphones, sound bars, and portable audio devices.
4. If these practices are followed for mixing, it is recommended to listen in a more “typical” listening space to evaluate translation to homes, where more spatial elements will be experienced. This should certainly be done at the mastering stage. Listeners often prefer some lateral reflections to compensate for limitations of stereo (see Section 4.3).

5. Art is being created, so use any available electronic processing of the program to get the desired "sound". Then explore how others might hear it.
    a. Listen at a variety of sound levels to evaluate bass balance, as it is affected by the equal-loudness contours (Section 9.4).
    b. Use a few high-pass filter settings to assess how the mix sounds with reduced bandwidth, different loudspeakers are not needed. This does not account for the fact that small bluetooth loudspeakers are active, with equalization and often a protection algorithm that turns down the bass as volume is increased. Uncertainties will exist. Complicating matters is the fact that even inexpensive headphones can have excellent bass extension and quality.
    c. Set up some preset equalizations to achieve frequency responses that assist in the tracking and mixing processes, but remove them for the final judgment. The sound of an NS-10M (or other favored near-field loudspeaker) is achievable by equalizing a neutral loudspeaker, and it can be switched off.
    d. If it is wished to explore spectral balances that portions of the public might be attracted to, use preset equalizations and see how the mix sounds. As has been pointed out several times in this book, consumer loudspeakers are no longer the random mix they used to be; there is a recognizable move to neutrality. Bass response remains a major uncontrolled factor as it is dominated by rooms.

The main caution to recording engineers is to avoid letting an idiosyncratic loudspeaker or room that you "like" to become part of the recording. Nobody else will ever hear it. The art is uniquely colored.

I once had a long discussion with an accomplished recording engineer who concentrated on the classical repertoire. I was told that a particular

loudspeaker, the B&W 802N, was particularly well suited for monitoring classical music. I happened to know the product. As described in Section 2.9.1 it is not a neutral loudspeaker, exhibiting problems in both measurements and in subjective evaluations. However, it flattered the sound of an orchestra miked from above, where excessive high frequencies are radiated. Instead of equalizing the mix, the loudspeakers were doing the equalization. The mediocre ratings in double-blind subjective testing tell us that the sound quality component of commercial recordings does not "translate" well through that loudspeaker. The evidence is in the anechoic measurements.

Opinions are always with us, as are biasing influences. Sighted biases are known, but there are more. The research described in Chapter 5 showed that listeners were able to separate the sound of the loudspeakers from that of the room in monophonic/hard-panned situations. Dramatically changing rooms did not significantly alter subjective opinions of loudspeakers auditioned within them in multiple-loudspeaker comparisons.

However, when stereo and multichannel recordings contain real or synthesized reflective/reverberant information about large spaces the situation changes. The multiple active channels reduce the ability of listeners to hear timbral details, specifically resonances, which are the sources of unwanted coloration in loudspeakers, as well as the desirable timbral components of voices and musical instruments. The perception of the large space in recordings dominates that of the listening room.

This is a disruptive observation in that it implies that the acoustical contributions of the listening room are diminished during multichannel activity. A spacious multichannel recording contributes greatly to the pleasure of listening, and in the process, it appears to exercise significant influence on what we are able to discern about the playback loudspeakers, the room, and very likely aspects of the program itself. Only in hard-panned, monophonic, components are our critical abilities restored. This is why it is necessary to evaluate loudspeakers in mono comparson tests, and enjoy them in stereo and multichannel systems.

12.5.5 An Interesting Experiment: Blackbird Studio C

Every so often someone does something that simply cannot be ignored, and this an example. George Massenburg, a noted recording engineer had the opportunity to create an ambitious new studio, and he teamed up with Peter D'Antonio who for decades had been a leader in scientifically based diffuser design and measurement. The result was Blackbird Studio C, a dramatic departure from traditional control room design.

Massenburg set the goals as a wide sweet spot, a high-definition soundstage and an acoustically immersive experience. These are subjectively judged, not acoustically measurable, objectives. They are also the desires of listeners at home. The requirement for "an acoustically immersive experience" is interesting, because the multichannel/immersive audio system should create the immersive experience, not the recording control room.

The room that was built had the entire walls and ceiling covered with up to 3 ft (1 m) deep broadband 2D diffusers. A visit to the website (www.blackbirdstudio.com; see Studio C) gives an indication of the complexity of the design. The deep, intricately fabricated wooden diffusers apparently weigh 100,000 pounds (45,360 kg). The reflected sound field is probably unique, because of the extremely effective dispersion of direct and reflected sounds throughout the room. Figure 12.28 shows the ETC measurement.

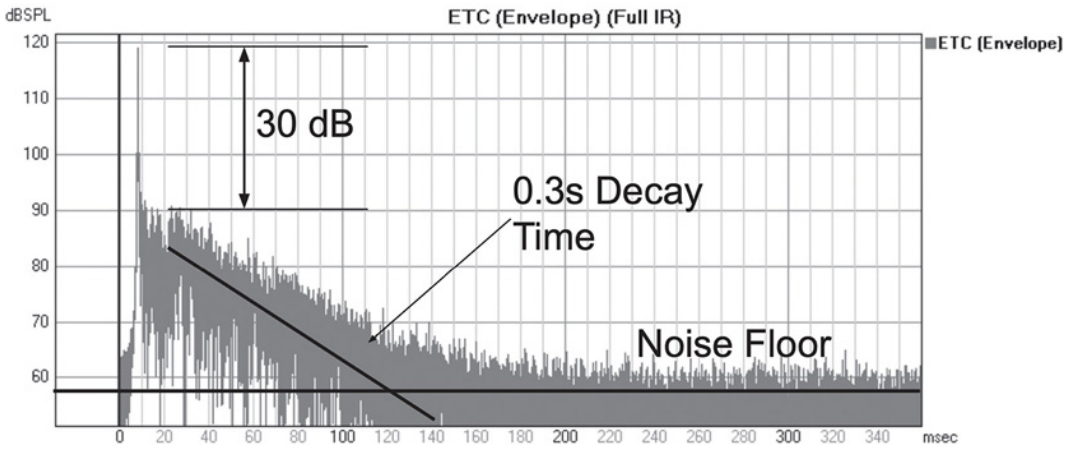


FIGURE 12.28 *The measured energy time curve (ETC) inside Blackbird Studio C. Courtesy of Peter D'Antonio.* ⏎

Whatever the expectations were, the reality is interesting. An immediate 30 dB drop is very like an anechoic space. The 0.3 s decay time (RT) that follows implies that the energy has been very uniformly dispersed in time and space. The high density of reflections from the diffusing surfaces that start almost immediately is the unique feature. The creators called it "Ambechoic"—an ambient anechoic space.

If multiple mini-reflections perceptually add up to more than their visual appearance, as discussed in Section 12.4.4, this room represents perhaps the most extreme place to test the theory. Apparently, it sounds more "live" than the ETC suggests, which supports the theory. Mixers reportedly adapt to its unique ambiance and are pleased with what they hear, and the recordings created in it. It is currently being used as an Atmos immersive music control room, which is not the most demanding of formats. The listening room tends to be less of a factor as the number of channels and recorded spatial information increases. It is also a successful space for performances, which is not surprising as a complex reflected sound field is a valuable asset.

Fazenda et al. (2015) say:

> The experience of this room is that one is unaware of sound reflection from the walls: it sounds almost anechoic, yet it has reverberation! Stereo and multi-channel material played in this room has images that are stable over a wide listening area.

It appears that the control room is embellishing the experience in ways that are appealing to recording engineers, and that would be appealing to customers of the recorded product. As it stands, only the people in that control room will ever have the experience. Building more elaborate physical structures is not the future. The challenge is to discover what it is that causes the desirable perceptions and find ways to deliver an

approximation of it through the multichannel audio system they are listening to. Then the experience can be shared.

Because the design objectives were subjective perceptions, it would be educational to have some binaurally recorded double-blind subjective comparisons of listening experiences in this and other environments. It must be difficult not to be influenced by the dramatically distinctive visual presentation of the room in sighted evaluations. The room definitely raises some interesting psychoacoustic questions and hints at some interesting signal processing developments. Until there are answers, there are opinions.

12.6 THE ACOUSTICAL TREATMENT OF LISTENING ROOMS

I have thought about this over the years and there is no solution to the design of interior acoustical treatment that can satisfy all users. Satisfaction is a multidimensional feeling and there are many levels of gratification, up to showplace custom home theaters where “gilt and glitter” are important ingredients.

The good news is that one can have impressive experiences in well-furnished typical rooms if the loudspeakers have been well chosen, and the bass issue—that everybody has—is addressed. Beyond that, enthusiasts may find ways to further improve the experience. A very common problem in small rooms is that listeners end up close to a wall behind them. If it is a reflective surface there is a problem (see Section 13.6). The solution is a resistive absorber, fiberglass or acoustic foam, not less than 4 inches thick placed behind and slightly above the listener’s head. It can be covered with acoustically transparent fabric, like loudspeaker grille cloth, or similar open weave material that is easy to blow through to make it more attractive. Diffusers are NOT recommended for this location as they deliver decorrelated sound to the ears degrading the imaging.

Figure 12.29 is intended to provide inspiration. It combines many ideas expressed in the book to this point, laid out for a large home theater. For smaller ones and stereo listening rooms, just scale appropriately.

Loudspeakers and acoustical treatments are sometimes not very attractive, and so it is common to stretch fabric over large portions of walls. This is not a problem if the fabric is acoustically transparent, but many are not, including some advertised as such, in which cases a lot of careful design, and the sound of the room, are degraded when the "finishing touch" is added.

The basic requirements can be summarized as follows:

All absorbers, wherever they are located, should be 3 to 4 inches (76 to 101 mm) thick if possible. Recall that as shown in Figure 12.20, a space behind an absorbing panel is an inexpensive way to improve performance. Filling the space with low-density fiberglass is beneficial.

The **floor** is covered with wall-to-wall clipped-pile carpet on felt underlay that is as thick as is practical. Felt acts as a fibrous absorber; most carpet underlays do not. The carpet needs to be porous, e.g. jute backed, not rubber backed.

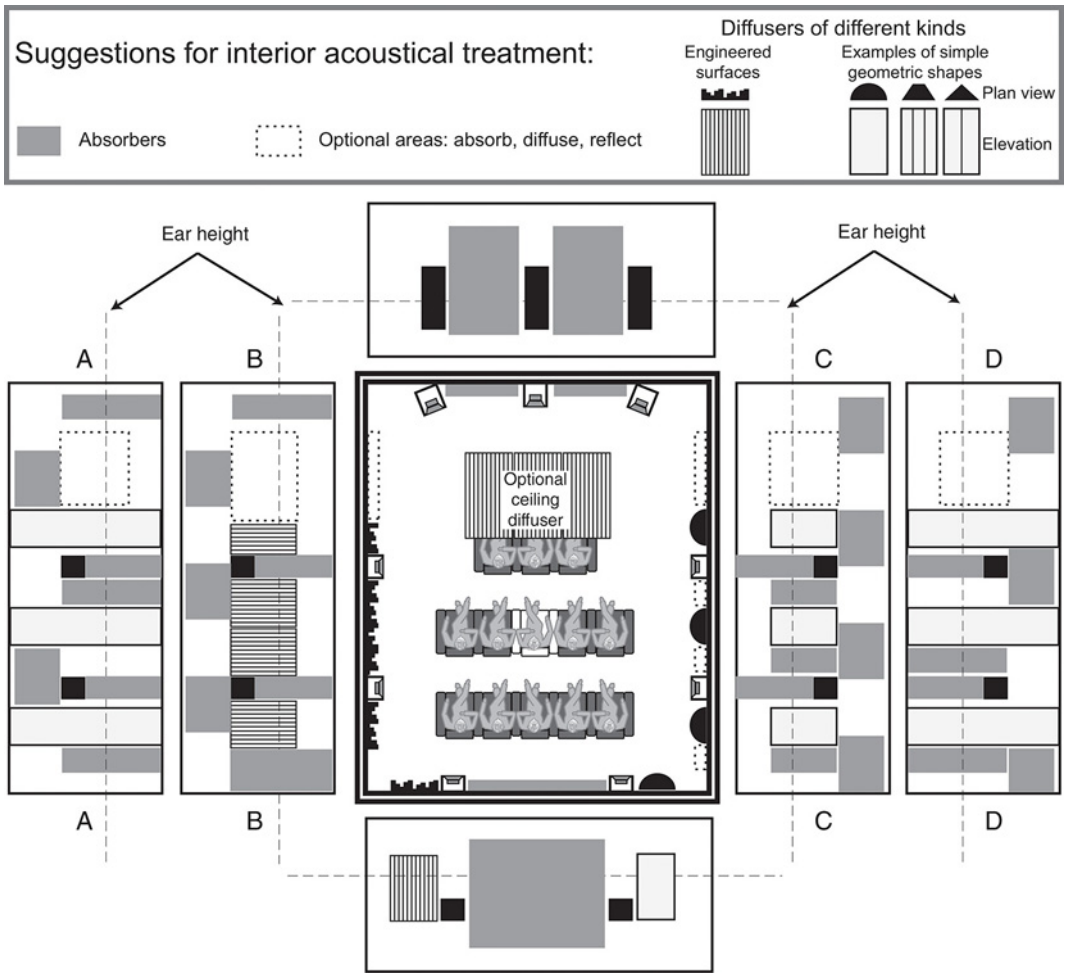


FIGURE 12.29 *This figure illustrates a few of the many possible ways to combine acoustical materials within a home theater. It shows a floor plan with walls folded down to show how materials might be arranged on them. The white seat in the middle is the prime listening location. Wall B shows a long array of engineered diffusers in a band around ear level. Wall C shows a version that uses hemi-cylindrical geometric shapes intermixed with reflecting surfaces and absorbing panels. Walls A and D show mirrored treatments in which the diffusing shapes have been extended floor to ceiling for visual effect. Absorbing panels have been placed in staggered locations to avoid flutter echoes. Many artistic variations are possible, and changes will be necessary to accommodate different numbers of surround channels. The*

dashed lines that identify the ear height of seated listeners must be adjusted to follow the floor contours of staged seating. ⏎

The locations of the **first side-wall reflections** at the front of the room are specified as areas for optional treatment. This is a topic that stimulates strong opinions and arguments among experienced listeners. Historically, absorbing side-wall reflections benefited badly designed loudspeakers, both consumer and professional. Recording control rooms often employed absorption and deflecting surfaces to eliminate early reflections (Section 3.5). Recording engineers often prefer to work in a dominant direct sound field. However, with loudspeakers exhibiting well-behaved off-axis performance, extensive listening tests among consumers revealed that there were benefits of early reflections in alleviating problems fundamental to stereo itself (Section 4.3.2). In stereo listening, "perfection" does not exist. Listeners preferring pinpoint imaging soundstage can choose to absorb or diffuse side-wall reflections. Those preferring a more open, spacious soundstage can leave those areas as flat wall surfaces and choose wide dispersion loudspeakers. The choice is influenced by music preferences. Hard-panned left and right sounds are essentially monophonic and usually benefit from some local reflections. When multiple channels are operating simultaneously, these reflections are swamped by the recorded spatial sounds and become neutral factors. With well-designed loudspeakers, having relatively uniform directivity, the effects of these side-wall reflections range from neutral to slightly beneficial, the choice can be left to the customer, there is no right or wrong.

The center portions of the **front and rear walls** are mostly absorbing, with scattering devices towards the sides of the rear wall. Avoid placing diffusers close behind listeners; the stereo imaging will be degraded.

The **side walls** are a mixture of reflection, absorption and scattering/ diffusing devices. Diffusers designed to scatter the sound horizontally should be located in the region about one foot below and about three feet above ear level.

Diffusers can be of the engineered-surface type, which are attached to the walls, or they can be simple geometrical shapes that can be attached to walls or integrated into structures, constructed on site using conventional building materials. To assist listeners in the perception of immersion or envelopment, engineered surfaces should be not less than about 8 inches (20.3 cm) deep, and of the type that disperses sound horizontally. Many curved and multifaceted convex geometric shapes will work, but hemi-cylinders are best. They should be not less than about 1 ft (30 cm) deep. For visual effect it is possible for the geometric shapes to extend floor to ceiling, imitating columns. Engineered diffusing surfaces can be arranged side by side to cover an area, as shown in B, but to work well geometric shapes must be separated as shown in A, C and D. Both types of diffusers can be mixed.

Absorbing material and blank reflecting areas should be arranged so that walls facing each other do not present opportunities for flutter echoes (in the drawing facing walls A and D, and B and C are designed with staggered areas of absorption and reflection). The total amount of absorbing material in the room must be sufficient to meet the reverberation time criteria.

Use geometry or a mirror to determine the best location for the optional diffuser on the **ceiling**. It should scatter sounds from the front loudspeakers that would normally be reflected to the head locations of audience members. This is most effective if it is an engineered surface designed to scatter sounds towards the sides of the room. Absorbing the ceiling reflection is a rational option.

The corners of the room are available for low-frequency absorbers. These are preferably of the membrane/diaphragmatic/panel type, because they are located in high-pressure regions of the low-frequency standing-wave patterns. Corners are also good locations for multiple subwoofers (Chapter 14).

12.6.1 Estimating Reverberation Time

In the context of home listening spaces, the correct RT is a generous target, ranging from “too dead” (below about 0.2 s) to “too live” (above about 0.5

s), both of which are easily recognizable by walking into the room and carrying on a conversation. That commentary might seem frivolous, but it is done to compensate for the sometimes overly fastidious attention devoted to RT (see Section 12.1). Except at bass frequencies, what is heard by listeners in small rooms is dominated by direct and early reflected sounds. Therefore, frequency-dependent absorption is a factor, but specific to the reflection points of those lateral early reflections related to the front channels that provide most of the important sounds, and these are related to the orientation and directivity of the loudspeakers (see Figure 8.7 and related discussion).

Let us compute where we stand using the room in the example shown in Figure 12.29, specifically using wall treatment configuration "B" (and its corresponding opposite with staggered areas of absorption) because it offers some interesting discussion. For this 20 × 24 × 9 ft (6.1 × 7.3 × 2.7 m) room, the Sabine reverberation time (Section 12.1.2) would be calculated (in imperial units) as RT = 0.049 V/A. V = 4320 ft^3, so the total amount of absorption (A in sabins) required in the room to achieve a midfrequency (say, 500 Hz) RT of 0.4 s is 530.

The carpet on the floor has an absorption coefficient of about 0.5, so we multiply that area by the absorption coefficient, giving us 20 × 24 × 0.5 = 240 sabins. The absorbing areas specified for the front and rear walls add up to about 106 ft^2 × 1.0 = 106 sabins (they are thick, resistive absorbers that are assumed to absorb perfectly at 500 Hz). Scattered patches of absorption on the side walls yield 110 sabins.

If RPG Modffusors (exposed absorption coefficient = 0.2, fabric covered it is around 0.6, but as shown in Figure 12.27 it is frequency dependent) were used along the side and rear wall, and they would contribute about 128 × 0.2 = 26 sabins, but if they were covered with fabric for appearance, that would increase to 77 sabins. Adding them up: 240 + 106 + 110 + 26 (77) = 482 (533)—close to the target of 530 sabins. The RT would be a maximum of 0.44 (0.4) s. The additional absorbing surface area of the seating and the people in the seats would further reduce the RT, as would the additive effects of scattering by the furniture and diffusers. If geometric or other

kinds of diffusers are used on the side walls, simply mix in enough pure absorption with the hard scattering shapes to approximate the numbers in the preceding prediction.

There is no single RT that guarantees satisfaction because the requirements for listening to stereo music is so different from watching immersive movies. RTs for music can drift towards 0.5 s, while those for movies drift towards 0.2 s. A safe middle ground is 0.3 s to 0.4 s, which embraces well-furnished rooms in many parts of the world, but there are regional differences.

This is my summarizing of what might be important. CEDIA/CTA—RP22 (2023) provides much information on designing custom immersive home theaters from the consultant's design/installation perspective.

REFERENCES

Angus, J. A. S. (1997). "Controlling Early Reflections Using Diffusion", *102nd Convention, Audio Eng. Soc.*, Preprint 4405. ⏎

Angus, J. A. S. (1999). "The Effects of Specular Versus Diffuse Reflections on the Frequency Response at the Listener", *106th Convention, Audio Eng. Soc*., Preprint 4938. ⏎

ANSI/CTA-2034-A (2015). "Standard Method of Measurement for In-Home Loudspeakers", *Consumer Technology Association, Technology and Standards Dept*. www.CTA.tech. ⏎

Arau-Puchades, H. (1988). "An Improved Reverberation Formula", *Acustica*, vol. 65, pp. 163–180. ⏎

Beranek, L. L. (1986). *Acoustics*, Acoustical Society of America, New York. ⏎

Bradley, J. S., Sato, H. and Picard, M. (2003). "On the Importance of Early Reflections for Speech in Rooms", *J. Acoust. Soc. Am*., vol. 113, pp. 3233–3244. http://dx.doi.org/10.1121/1.1570439 ⏎

Bradley, J. S., Soulodre, G. A. and Norcross, S. (1997). "Factors Influencing the Perception of Bass", *J. Acoust. Soc. Am*, vol. 101, p. 3135. http://dx.doi.org/10.1121/1.419017 ⏎

CEDIA/CTA—RP22 (2023). “Immersive Audio Design Recommended Practice”. www.cta.tech ⏎

Cox, T., Dalenbäck, B-I., D’Antonio, P., Embrechts, J., Jeon, J., Mommertz, E. and Vorlander, M. (2006). “A Tutorial on Scattering and Diffusion Coefficients for Room Acoustic Surfaces”, *Acta Acustica United with Acustica*, vol. 92, pp. 1–15. ⏎

Cox, T. and D’Antonio, P. (2004). *Acoustic Absorbers and Diffusers*, Spon Press, London & New York. ⏎

Cox, T. and D’Antonio, P. (2009). *Acoustic Absorbers and Diffusers*, 2nd edition, Taylor & Francis, Abingdon, Oxon, England. ⏎

Cremer, L. and Müller, H. A. (translation by T. J. Schultz) (1982). *Principles and Applications of Room Acoustics*, Vol. 1 & 2, Applied Science Publishers, London. ⏎

Cuharero, J., Hänninen, T. and Kokki, T. (2020). “Angle-Dependent Absorption of Sound on Porous Materials”, *Acoustics*, vol. 2, pp. 753–765. ⏎

Dalenbäck, B.-I. (2000). “Reverberation Time, Diffuse Reflection, Sabine and Computerized Prediction”. www.rpginc.com/research/reverb01.htm ⏎

Devantier, A. (2002). “Characterizing the Amplitude Response of Loudspeaker Systems”, *113th Convention, Audio Eng. Soc*., Preprint 5638. ⏎

Fazenda, B., Angus, J. and Cox, T. (2015). “The SOS Guide to Control Room Design”. www.soundonsound.com/techniques/sos-guide-control-room-design. Accessed October 2024. ⏎

Fitzroy, D. (1959). “Reverberation Formula Which Seems to Be More Accurate with Nonuniform Distribution of Absorption”, *J. Acoust. Soc. Am*., vol. 31, pp. 893–897. ⏎

Gedemer, L. (2015). “Predicting the In-Room Response of Cinemas from Anechoic Loudspeaker Data”, *Audio Eng. Soc. 57th Conference: The Future of Audio Entertainment Technology*. ⏎

Gilford, C. L. S. (1959). “The Acoustic Design of Talks Studios and Listening Rooms”, *Proc. Inst. of Electrical Eng*., vol. 106, pp. 245–258.

Reprinted in *J. Audio Eng. Soc*., vol. 27, pp. 17–31, 1979.

Green, R. and Holman, T. (2010). "First Results from a Large-Scale Measurement Program for Home Theaters", *Audio Eng. Soc. 129th Convention*, Preprint 8310.

Griffin, J. R. (2003). "Design Guidelines for Practical Near Field Line Arrays". www.audiodiycentral.com/resource/pdf/nflawp.pdf

Hughes, R., Cox, T., Shirley, B. and Power, P. (2016). "The Room-in-room Effect and Its Influence on Perceived Room Size in Spatial Audio Reproduction", *141st Convention Audio Eng. Soc*., Preprint 9621.

IEC 60268-13 (1998). Technical Report—Type 3, "Sound System Equipment—Part 13: Listening Tests on Loudspeakers", *International Electrotechnical Commission*, Geneva, Switzerland.

ITU-R BS.1116-3 (2015). "Methods for the Subjective Assessment of Small Impairments in Audio Systems", *International Telecommunication Union*, Geneva.

Keele, D. B. and Button, D. J. (2005). "Ground-plane Constant Beamwidth Transducer (CBT) Loudspeaker Circular-arc Line Arrays", *119th Convention, Audio Eng. Soc*., Preprint 6594.

King, R., Theriault, M. and Massenburg, G. (2023). "A Practical Approach to the Use of Center Channel in Immersive Music Production", *155th Convention Audio Eng. Soc.*, Paper 168.

Leembruggen, G. (2015). "Low Frequency Issues: Cinema/Home/Internet", *Audio Eng. Soc., 57th Conference*, "The Future of Audio Entertainment Technology: Cinema, Television and the Internet".

Linkwitz, S. (2009). "The Challenge to Find the Optimum Radiation Pattern and Placement of Stereo Loudspeakers in a Room for the Creation of Phantom Sources and Simultaneous Masking of Real Sources", *127th Convention Audio Eng. Soc*., Paper 7959.

Lipshitz, S. and Vanderkooy, J. (1986). "The Acoustic Radiation of Line Sources of Finite Length", *81st Convention, Audio Eng. Soc*., Preprint 2417.

Meyer, Jürgen (2009). *Acoustics and the Performance of Music*, Fifth Ed., Springer.

Newell, P., Santos-Dominguez, D., Torres-Guijarro, S. and Castro, S. (2015). “The Relationship Between Subjective and Objective Response Differences at Different Heights above Cinema Seating”, *Proc. Inst. of Acoust*., vol. 37, pt. 4. ⏎

Olive, S. E., Welti, T. and McMullin, E. (2013). “Listener Preference for In-Room Loudspeaker and Headphone Target Responses”, *135th Convention, Audio Eng. Soc*., Paper 8994. ⏎

Rathsam, J. and D‘Antonio, P. (2006). “Reflectivity of Absorptive Surfaces”, *4th Joint Meeting of the Acoust. Soc. Am. and the Acoust. Soc. Japan*, Honolulu, HA, November. ⏎

Robinson, P., Walther, A., Faller, C. and Braasch, J. (2013). “Echo Thresholds for Reflections from Acoustically Diffusive Architectural Surfaces”, *J. Acoust. Soc. Am*., vol. 134, pp. 2755–2764. ⏎

Schultz, T. J. (1983). “Improved Relationship between Sound Power Level and Sound Pressure Level in Domestic and Office Spaces”, *Report No. 5290, Am. Soc. of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE)*, prepared by Bolt, Beranek and Newman, Inc. ⏎

Silzle, A., Geyersberger, S. Brohasga, G., Weninger, D. and Leistner, M. (2009). “Vision and Technique behind the New Studios and Listening Rooms of the Fraunhofer IIS Audio Laboratory”, *Audio Eng. Soc. 126th Convention*, Preprint 7672. ⏎

Smith, D. L. (1997). “Discrete-element Line Arrays—Their Modeling and Optimization”, *J. Audio Eng. Soc*., vol. 45, pp. 949–964. ⏎

SMPTE TC-25CSS (2014). “B-Chain Frequency and Temporal Response Analysis of Theaters and Dubbing Stages”, *Soc. Motion Picture and Television Eng*. ⏎

Toole, F. E. (1986). “Loudspeaker Measurements and Their Relationship to Listener Preferences”, *J. Audio Eng. Soc*., vol. 34, pt. 1, pp. 227–235, pt. 2, pp. 323–348. ⏎

Toole, F. E. (2006). “Loudspeakers and Rooms for Sound Reproduction—A Scientific Review”, *J. Audio Eng. Soc*., vol. 54, pp. 451−476. ⏎

Toole, F. E. (2015). “The Measurement and Calibration of Sound Reproducing Systems”, *J. Audio Eng. Soc*., vol. 63, pp. 512–541. This is

an open-access paper available to non-members at www.aes.org www.aes.org/e-lib/browse.cfm?elib=17839 ⏎

Toole, F. E. and Olive, S. E. (1994). “Hearing Is Believing vs. Believing Is Hearing: Blind vs. Sighted Listening Tests and Other Interesting Things”, *97th Convention, Audio Eng. Soc*., Preprint 3894. ⏎

Zacharov, N., Bech, S. and Suokuisma, P. (1998). “Multichannel Level Alignment, Part II: The Influence of Signals and Loudspeaker Placement”, *Audio Eng. Soc. 105th Convention*, Preprint 4816. ⏎

CHAPTE
R 13

Acoustics: Adjacent-Boundary Effects

Floyd Toole

DOI: 10.4324/9781003477495-14

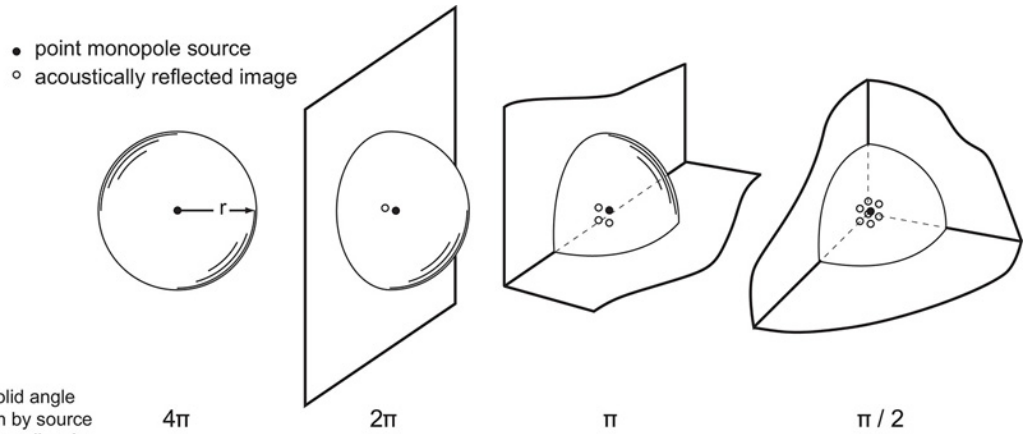
13.1 THE EFFECTS OF SOLID ANGLES ON SOUND RADIATED BY POINT SOURCES

At 20 Hz, the wavelength is 56.5 feet (17.25 m). At this and similar frequencies any practical separation between a loudspeaker and a room boundary is "small". In addition, any practical loudspeaker will be small relative to the wavelength, and therefore it will radiate sound in an omnidirectional manner. For both of these reasons the conditions are met for the classic set of relationships shown in Figure 13.1.

In Figure 13.1, the technical description of full spherical omnidirectional radiation is that the sound source "sees" a solid angle of 4π steradians. It is a "free field" with no surfaces to reflect or redirect the radiated sound. Placing the sound source on or in a large plane surface reduces the solid angle into which the sound radiates by half, to 2π steradians. Energy that would have traveled into the rear hemisphere is reflected forward; there is a

reflected acoustical “image” of the source. Additional surfaces, positioned at right angles, reduce the solid angle by half, to π steradians, and then to π/2 steradians. The number of reflected images increases correspondingly.

It is commonly heard in the audio industry that each factor-of-two reduction of the solid angle increases the sound level by 3 dB. The truth is that this may or may not be the case, depending on circumstances. As seen in Figure 13.1, sound pressure level increases in 6 dB increments, but sound *power* (the total sound energy radiated into and distributed over the solid angle) goes up in increments of 3 dB. It can be seen that the sound pressure level, measured at a constant distance from the sound source *in these otherwise reflection-free circumstances*, goes up by 6 dB for each halving of the solid angle. Strictly this applies only to the direct sound. In a large reflective venue the reflected sounds represent the total power radiated at all angles, and in the incoherent summation at a distant point the gain will be closer to the 3 dB per halving of the solid angle.



Solid angle seen by source (steradians)	4π	2π	π	$\pi / 2$
Sound pressure at radius r	p_r (0 dB)	$2p_r$ (+6 dB)	$4p_r$ (+12 dB)	$8p_r$ (+18 dB)
Sound intensity at radius r	I_r (0 dB)	$4I_r$ (+6 dB)	$16I_r$ (+12 dB)	$64I_r$ (+18 dB)
Surface area at radius r	A_r	$\frac{A_r}{2}$	$\frac{A_r}{4}$	$\frac{A_r}{8}$
Sound power radiated into the solid angle (power = IA)	$= I_r A_r$ $= P_r$ (0 dB)	$= 4\ I_r \frac{A_r}{2}$ $= 2P_r$ (+3 dB)	$= 16\ I_r \frac{A_r}{4}$ $= 4P_r$ (+6 dB)	$= 64\ I_r \frac{A_r}{8}$ $= 8P_r$ (+9 dB)

FIGURE 13.1 *Various measurable quantities at a fixed distance from a physically small omnidirectional sound source (a point monopole) in several basic locations closely adjacent to large flat surfaces. The sound in this illustration radiates into anechoic space. In normal rooms these concepts apply only to the direct sound.* ⏎

But there is more to be aware of. What is shown in Figure 13.1 happens only at wavelengths that are long compared to the source size and the separation distance. With respect to where we place loudspeakers or listeners relative to the boundaries of a room, this normally applies only at very low frequencies. How low?

Practical measurements confirm the theory. Figure 13.2 shows measurements made at the NRCC in Ottawa, in the mid 1980s, in an anechoic chamber, and outdoors, in a parking lot adjacent to an isolated

flat-sided building. The only reflecting surfaces were the adjacent boundaries. The loudspeaker was a conventional 12-inch (305 mm) driver in a closed box, about 17 inches (432 mm) on a side. Noise problems prevented the acquisition of data below about 35 Hz, and a technical problem resulted in the corruption of the 2π steradian data, but the trend is very evident.

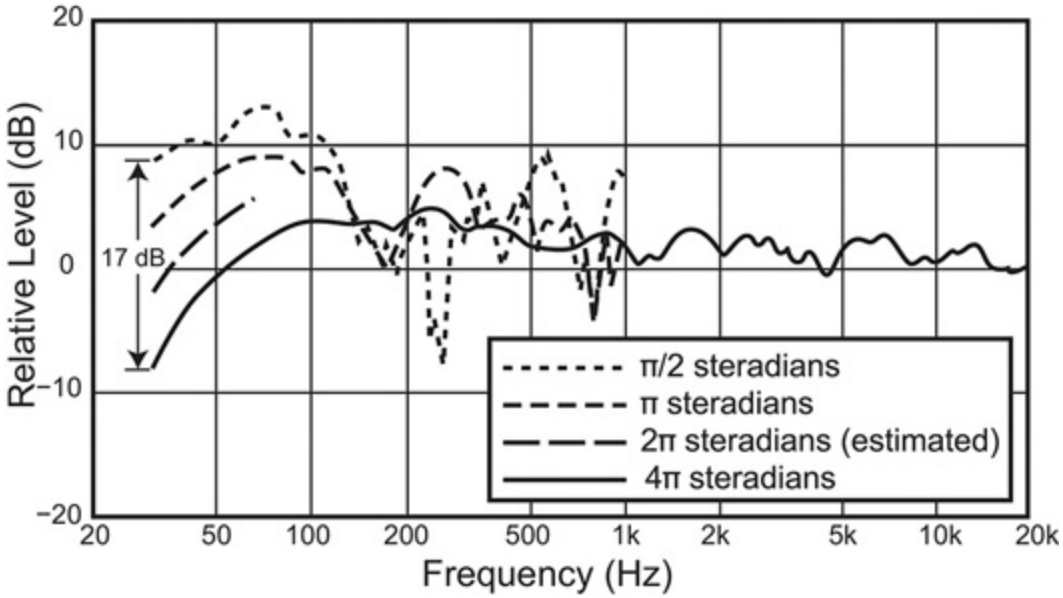


FIGURE 13.2 *A loudspeaker measured when it radiates into different solid angles as approximated by an anechoic chamber (calibrated at low frequencies) for the 4π condition, and by an outdoor parking lot adjacent to a rectangular building. All measurements were made at 2 m.* ⏎

First, at the lowest frequency, the curves are separated by very close to 6 dB for each halving of the solid angle, as Figure 13.1 predicts. The fact that there is a tiny discrepancy can easily be accounted for by the fact that the large powerful woofer is not exactly a vanishingly small "point monopole", it is not located "in" the wall/floor, and the building was sheathed in corrugated metal siding, which is not a perfect reflector. In any event, the approximate 6 dB difference begins to diminish immediately and just below 200 Hz all of the curves converge in a null. The reason is that the woofer

was facing forward, with the rear of the enclosure against the wall. This separation allowed the sound to travel backwards from the driver to the wall, and back again toward the driver where it interacted with sound being radiated by the diaphragm. The round-trip distance, about 36 inches (914 mm), is one-half wavelength, at 188 Hz, which is where the destructive acoustical interference occurs. This figure demonstrates the adjacent-boundary effect and its wavelength dependence. The very low-frequency sound level gain is seen, as well as a reduction in sound output because the sound source is at a finite distance from the reflecting surfaces.

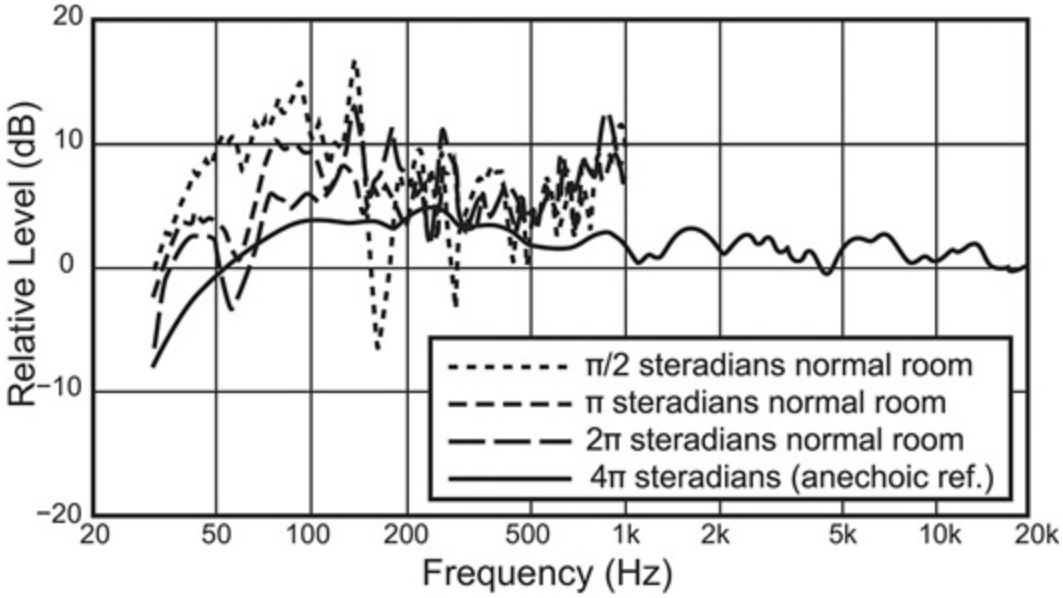


FIGURE 13.3 *The averages of four to six measurements made in the same loudspeaker/microphone relationship used in the previous figure, but moved to different locations within a normal listening room.* ⏎

What happens in a normally reflective small room? Figure 13.3 shows measurements made using the same loudspeaker in the same relationship to the microphone but positioned in the middle of the room (2□), on the floor against a wall (□), and in a corner (□/2). It was done multiple times in different locations around the room, and the results averaged—individual curves were heavily influenced by standing waves and were variable.

However, there is evidence of the microphone being near the null in the first-order vertical room mode at 62 Hz because this was common to all measurements. Otherwise, the untidy averaged curves may be impossible to interpret, but it is clear that there are gains in sound level for each halving of the solid angle. The phenomena that matter at very low frequencies need to be addressed for specific loudspeaker locations, addressing specific listener locations, as explained in Chapter 14.

13.2 CLASSIC ADJACENT-BOUNDARY EFFECTS

Figure 13.2 shows an acoustical interference dip at 188 Hz, caused by reflected sound from an adjacent boundary, the wall behind the loudspeaker in this case, destructively interfering with the direct sound radiated by the loudspeaker—the loudspeaker is omnidirectional at these frequencies. It shows up clearly in these outdoor measurements with only adjacent boundaries. In rooms there are multiple boundaries, some adjacent to and none very far away from the loudspeaker, each contributing something to variations in frequency response at listening locations.

The adjacent boundary phenomenon was known to acousticians (Waterhouse, 1958), but it was Allison and Berkovitz (1972) and Allison (1974) who brought it into the consciousness of audio people in papers that described, in measurements, the significant dimensions of the problem. It is sometimes known as the “Allison effect”.

Allison (1974, 1975) presents many examples of room curve shapes resulting from different arrangements of loudspeakers and corner boundaries. The curves are different from each other, depending on the distances of the woofers to the floor and each of the nearby walls. An average of 22 of these, in eight rooms, is shown in Figure 13.4. To get an impression of how this curve relates to anechoic measurements, the figure also shows 2π and 4π anechoic curves. The precise vertical alignment of these curves is uncertain, so focus on the shapes. It seems that the anechoic facilities were not calibrated at very low frequencies.

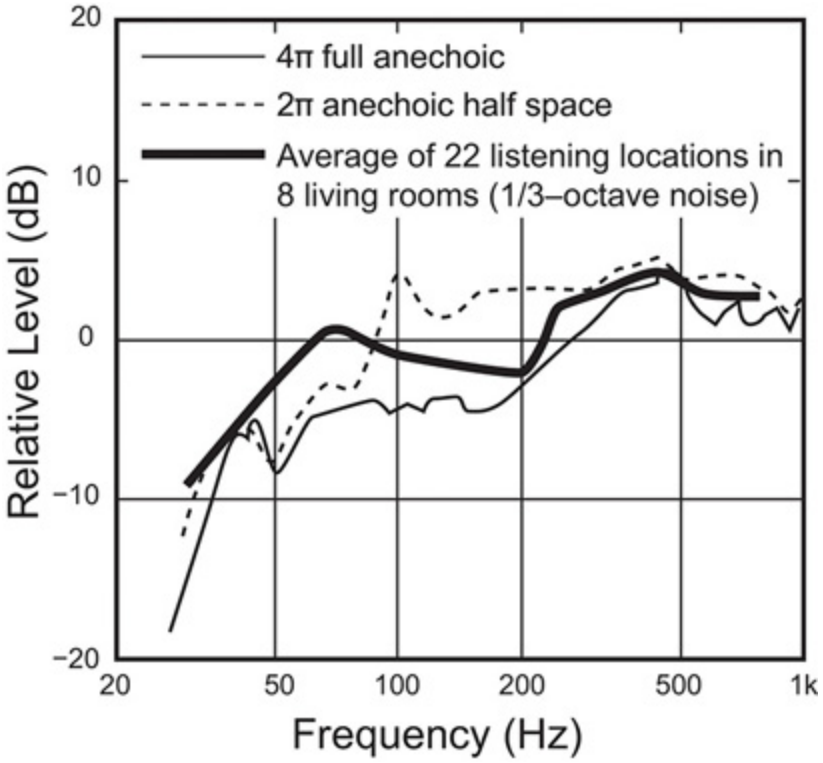


FIGURE 13.4 *A comparison of 2π and 4π anechoic measurements on an Acoustic Research AR-3a compared with the average of 1/3-octave measurements made at 22 listening locations in eight living rooms. Anechoic data from Allison and Berkovitz (1972, figures 4 and 9). The room data was also in this paper, but was more conveniently presented in Allison (1974, 1975).* ⏎

Each of the measured rooms would have had a distinctive signature superimposed on the low-frequency roll-off characteristics of the loudspeaker. Although there are uncertainties in the data, it is evident that moving from anechoic 4□ to 2□ and then to rooms, the very low bass shows progressive gain, but in the rooms, there is an irregular dip in output between about 80 Hz and 200 Hz attributable to adjacent boundary effects that are averaged over all measurements in the eight rooms. This is the adjacent boundary problem that needs to be addressed, but it needs to be addressed for each specific room/loudspeaker arrangement.

All of this is generally predictable for a specific situation, and Figure 13.5 shows the result of computer modeling of the interaction between a loudspeaker and a specific relationship to adjacent boundaries. Obviously, different loudspeaker locations would result in different results. In the model the loudspeaker has a perfectly flat frequency response, no low-frequency roll-off, so what is seen in the computed result is due entirely to the loudspeaker/boundary interaction.

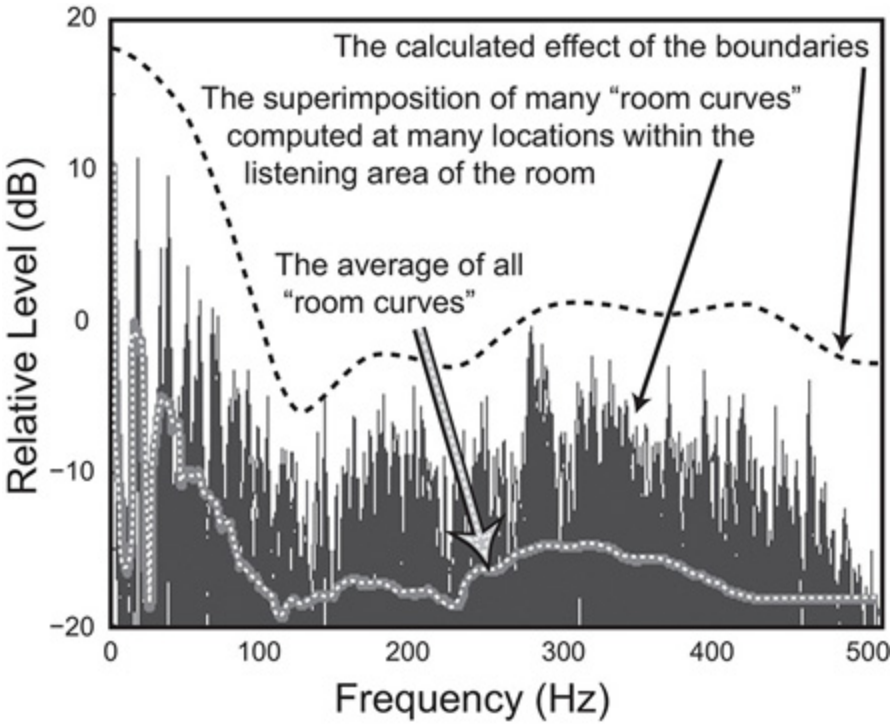


FIGURE 13.5 *A computer simulation of an omnidirectional loudspeaker situated in a "normal" relationship to the floor, ceiling and wall boundaries, in a "normal" listening room. Simulation by Todd Welti, Harman International.*

In Figure 13.5 the dotted curve shows the predicted effect of the adjacent boundaries on in-room frequency responses. To assist in understanding what is happening, the figure shows a superimposition of many steady-state frequency responses calculated for many different locations within the listening area. Each one is different because of standing waves and reflected sounds. Also shown is the average of all of these measurements. The

modeled loudspeaker had a flat frequency response, and here we can see the substantial room gain at very low frequencies (in contrast to the bass roll-off seen in Figure 13.4). This is, therefore, an estimate of the effect of the adjacent boundaries on the radiated sound power, *which is not specific to the sound arriving at any single listener location*.

The effects of room boundary interactions are in all room curve measurements, but in any individual one they may be obscured by other factors, such as standing waves. However, by averaging several room curves, measured at different locations—a spatial average—the effects of the position-dependent variations are reduced, and evidence of the underlying adjacent-boundary effects is more clearly seen. The average of the room curves is obviously similar in shape to the predicted curve.

This tells us that eliminating the adjacent-boundary effect will not eliminate all problems, but it is definitely one of the problems. The adjacent-boundary effect is one that, because it changes the acoustical radiation resistance experienced by the loudspeaker, the sound power radiated by the loudspeaker at different frequencies is altered. How can such a problem be addressed?

13.3 ALLEVIATING ADJACENT-BOUNDARY EFFECTS

13.3.1 Positioning the Woofer

The approach offered by Allison (1974, 1975) and Ballagh (1983) involves choosing the position of the loudspeaker with respect to the boundaries in a manner that minimizes the variations in frequency response at the listening locations. Obviously, the goal is to arrange for the woofer-to-boundary distances, floor and walls (length, width and height), to be as different as possible, while not being multiples of each other. The more different these distances are, the less will be the effect at any frequency, but there will still be an effect. Finding an optimum result may involve some trial and error. It may also result in visually unappealing, asymmetrical, or incorrect (in terms of stereo or multichannel imaging) locations for the loudspeakers.

Absorbing the boundary reflections is an option, but to be effective at these low frequencies, one is looking at large, thick devices beside and behind the loudspeakers, and treating the floor would not be practical.

13.3.2 Equalization

Positioning is a no-cost remedy that reduces but does not eliminate the problem. Equalization is another option, used in addition to or instead of positioning. The attraction of equalization is that it allows the loudspeakers to be located according to other criteria, and then the performance is electronically optimized.

Figure 13.5 shows that averaging several measurements within the listening area can reveal the underlying shape of the room-boundary effects, and thereby provide a basis for correcting the frequency response to meet whatever low-frequency target curve is decided on. However, there is another method, described by Pedersen (2003), in which the acoustic power output of the loudspeaker is evaluated in situ. The appropriate equalization correction to the frequency response is then applied to achieve a more uniform output.

Figure 13.6 shows a comparison of the two methods in the same room, one the result of making frequency response measurements at nine very different listening locations, and the other the result of a measurement of the sound power radiated by the loudspeaker. They are remarkably similar and, if any amount of spectral smoothing were incorporated, they would be even closer than shown. Obviously, this means that an in-room measurement of radiated sound power can identify the adjacent-boundary problems, and so can a spatial average of in-room measurements. Both methods allow us to separate out the adjacent-boundary problems, but the solution applies everywhere throughout the volume of the room, which is unlikely to be what is needed for any particular listening location. The effects of room resonances are added to this, so there will be seat-to-seat variations. If the equalization is performed at a single listening location, it will obviously be correct for that seat, and it will be a combination of adjacent boundary and room resonance problems.

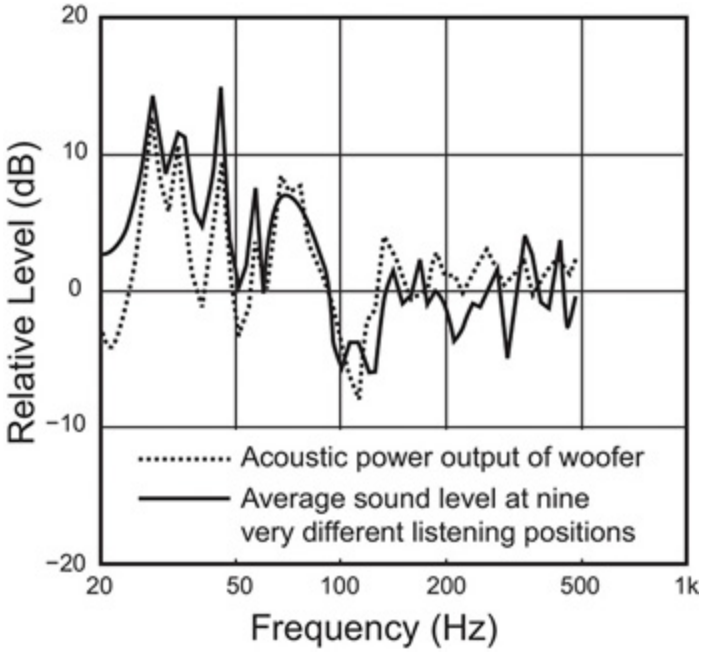


FIGURE 13.6 *A comparison of two methods of evaluating adjacent-boundary effects: a measure of the sound power radiated by the loudspeaker and the average of several room curves measured at different locations in the listening area. Redrawn from Pedersen (2003).* ⏎

Angus (2010) offers suggestions for using loudspeaker bass alignments to compensate for the large-scale bass effects of nearby boundaries. Small undulations and standing waves were not included.

The following chapter addresses the dominant problem of room resonances in the subwoofer frequency and range, and a convincing argument will be made for equalization of the low-bass subwoofer frequencies, including methods to diminish seat-to-seat variations. Here we see that in the frequencies above the subwoofer range, through the transition frequency range, equalization again is beneficial.

13.4 LOUDSPEAKER MOUNTING OPTIONS AND EFFECTS

In general, conventional loudspeakers should be used “free-standing”, away from walls. So-called bookshelf loudspeakers, presumably so named because they can fit a (sometimes large) bookshelf, are as often as not used on stands to bring the tweeters to seated ear level. The engineers designing these products cannot anticipate the details of any other manner of use, and so such loudspeakers are generally optimized to sound good in a free-standing mode. There are exceptions, of course: some loudspeakers are optimized to be mounted on walls or ceilings, and others are optimized for in-wall or in-ceiling mounting. These specialized products are clearly identifiable. It is instructive examine what happens to the sound radiated by a small bookshelf loudspeaker if we violate the rules and do not use it in the intended manner. Such a test took some preparation.

The chamber used for the following measurements was chosen because it could accommodate the apparatus, an 8-foot (2.44 m) square section of “domestic frame wall” on and into which loudspeakers were mounted. This chamber was not perfectly anechoic at the lowest frequencies and, although it had been calibrated for the measurement of loudspeakers in isolation, the presence of the massive wall structure introduced errors. Consequently, no claim is made for absolute accuracy at low frequencies in the measurements that follow, but they should be reliable in a comparative sense.

The tests began with an Infinity Primus 160 in the free field, an anechoic chamber, shown in Figure 13.7(a). The acoustical measurements shown are components of the spinorama: the conventional on-axis frequency response, measured at 2 m, the sound power output, and the directivity index, the difference between the two. Directivity index (DI) is 0 for an omnidirectional source, and here we see that this little 6.5-inch (165 mm) woofer approximates that quite well up to about 150 Hz. One can see the directionality progressively increase to about 7 dB at 2 kHz, above which the crossover network progressively attenuates the woofer and the tweeter takes over. Being small, the tweeter exhibits much better dispersion, about 4.5 dB around 4 kHz, and then it too becomes progressively more directional, reaching a DI around 9 dB at the highest frequencies. Just for

perspective, this is extremely good performance for an inexpensive product, listing at $220/pair at the time.

Figure 13.7(b) shows what happens when it is then mounted in a wall, with its front face flush with the surface. This is the classic 2π, half-space condition, which happens to be met by all in-wall and in-ceiling loudspeakers. What happens?

First, the bass increases, exactly as predicted by Figure 13.1. The sound pressure level goes up by roughly 6 dB (remember the measurements include some errors). Around 100 Hz it can be seen that in (a) the on-axis curve is about 3 dB below the 80 dB line, and in (b) it is about 3 dB above the line, for an increase in sound pressure level of about 6 dB. The sound power increases by about 3 dB in going from the 4π to the 2π conditions, again as predicted in Figure 13.1. The gains in acoustic output drop at higher frequencies because, as seen in (a), the loudspeaker is no longer perfectly omnidirectional; more of the sound is being radiated forward and not being reflected by the boundary. The DI at low frequencies is, as theory would predict, 3 dB for a half space.

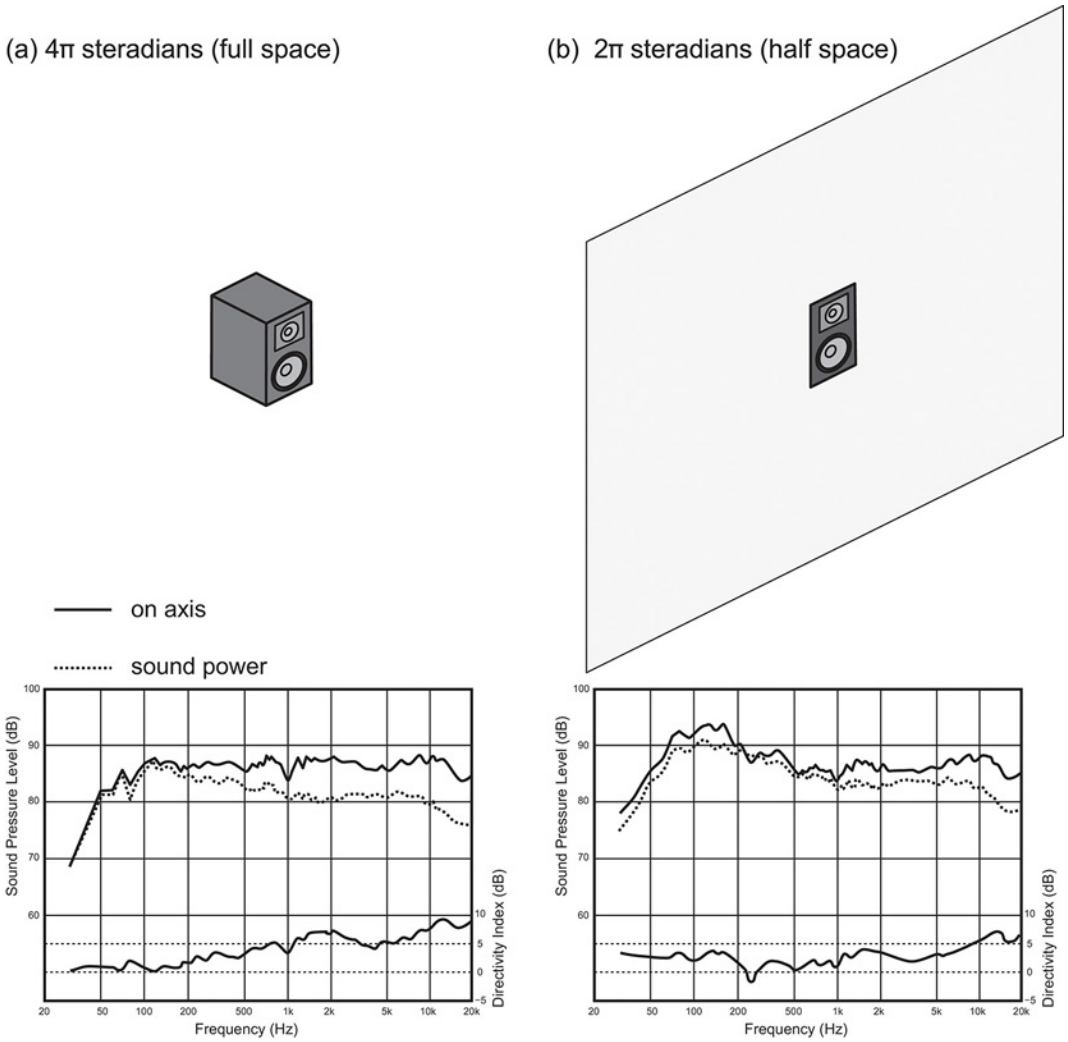


FIGURE 13.7 *The same small bookshelf loudspeaker measured in 4π, full anechoic, conditions and in 2π, half-space, conditions. The enclosure was carefully flush mounted.*

The overall conclusion is that mounting this excellent little loudspeaker in a wall has left its overall performance substantially intact, but the bass output has been greatly increased, making it sound fat, thick and tubby. After all, it was not designed to be used in this manner. The solution in this case is to turn the bass down. Any competent equalizer can do it, or the old-fashioned bass control may just be optimum if the "hinge" frequency is around 500 Hz.

The proper solution, if a loudspeaker is to be used in this manner, is to design it from the outset so that it has a flat axial frequency response when it is mounted in a wall. All in-wall and in-ceiling loudspeakers *should* be designed in this manner, but not all are.

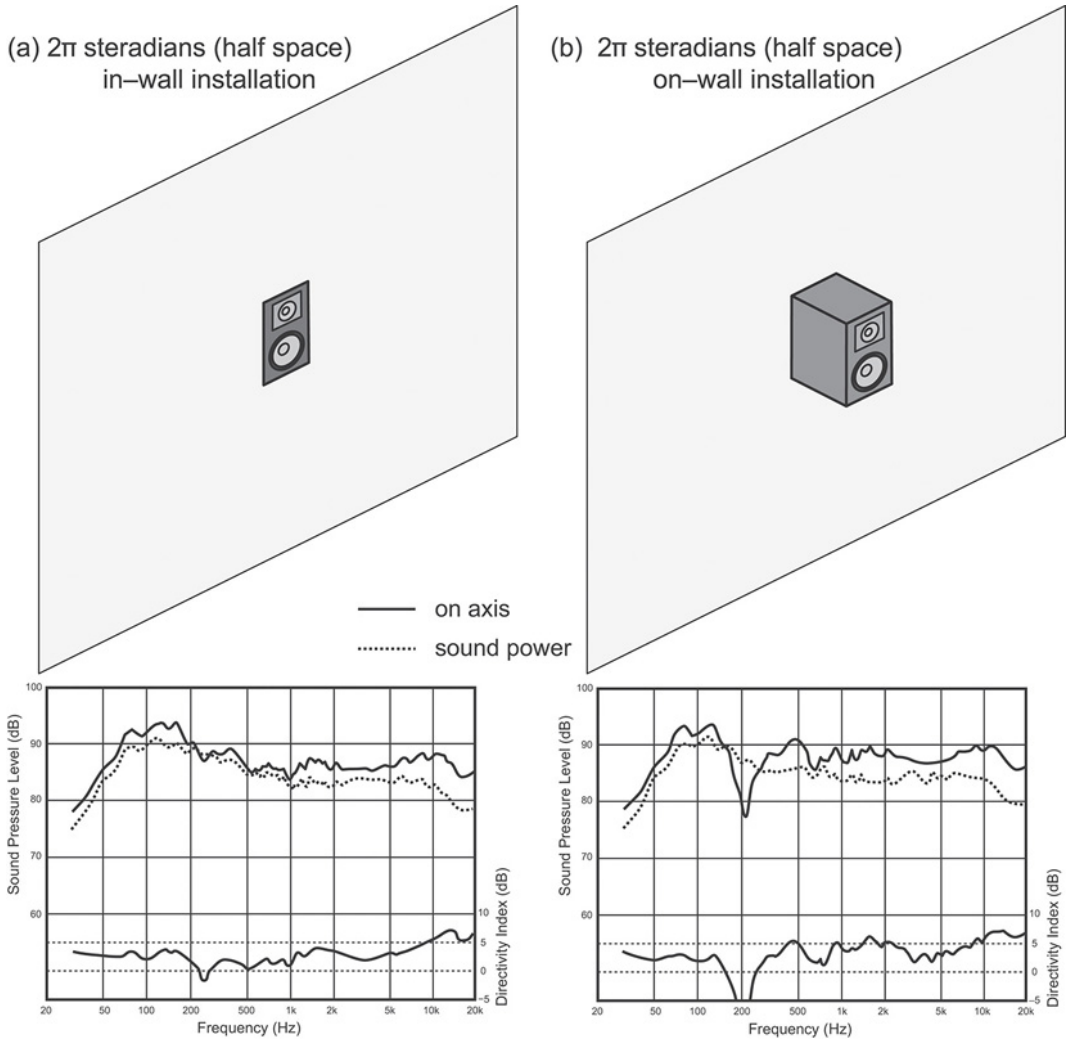


FIGURE 13.8 *A comparison of (a) in-wall and (b) on-wall placement of the bookshelf loudspeaker.*

Moving on, Figure 13.8(b) shows what happens when the loudspeaker is simply mounted on the surface of a wall. For comparison purposes the half-space data from Figure 13.7 are repeated in (a). It can be seen that there is a strong acoustical interference dip at about 220 Hz. We saw this before, back in Figure 13.2, but at a lower frequency. This loudspeaker is smaller, and

the round-trip distance from the woofer to the wall and back is shorter: about 31 inches (787 mm), which is one-half wavelength at about 220 Hz, the destructive-interference condition representing the first “tooth” in a comb filter. There is a hint of a second tooth, a partial cancellation, in the on-axis curve at 660 Hz, but one can presume that increasing source directivity at higher frequencies eliminates any higher-order cancellations. But, why is there no corresponding “hole” in the sound power curve? It turns out that there is, but it is not so easy to see as the dramatic event in the on-axis curve. This will be explained in the next section.

We have all seen it: loudspeakers sitting in otherwise empty cavities in bookcases, entertainment furniture and, embarrassingly, in expensive custom installations. Figure 13.9 shows what can result: in this case a perfectly good little loudspeaker has been seriously compromised by the installation. There is evidence of high-Q cavity resonances and diffraction effects created by the edges of the cavity. The everyday remedy of filling the cavity with absorbing material helps, but the root problem is still in evidence. Figures 13.7(b) and 13.8(a) show the much improved performance in a mounting where the cavity openings have simply been closed off with hard material.

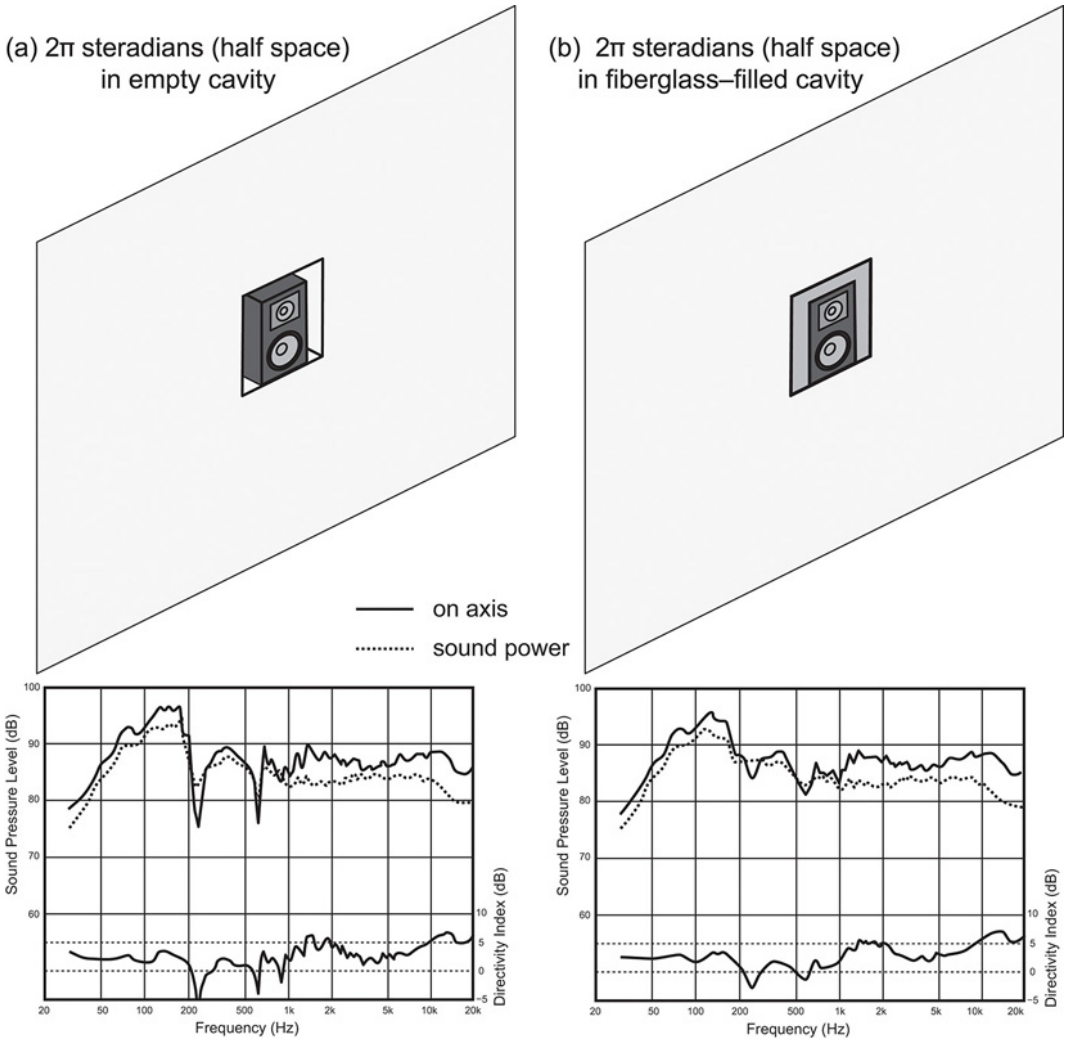


FIGURE 13.9 *A comparison of acoustical performance when the same small bookshelf loudspeaker is placed in a cavity—a bookshelf perhaps—as often happens in "entertainment system" furniture. In (a) the cavity is empty. In (b) the cavity has been filled with fiberglass.*

Reflecting on what has just been discussed, it can be concluded that there are really only two locations in which a loudspeaker has the potential of performing at its best: free standing, or flush mounted in a wall (or ceiling). All other options involve compromises.

13.4.1 An Example of Adjacent-Boundary Interference

Figure 13.8(b) showed an obvious acoustical cancellation dip in the on-axis frequency response, but not in the sound power. Figure 13.10 explains why.

The ray diagram in Figure 13.10 illustrates that the direct and reflected path length difference is maximum at 0° (on axis), trending to zero at 90° off axis. Therefore, the frequency at which the first destructive acoustical interference occurs rises from the 200 Hz dip seen in Figure 13.8(b) to ever increasing frequencies as the measurement point moves away from the forward axis. The declining off-axis radiation from the loudspeaker is a major factor in the progressive reduction in the cancellation dip. In the sound power spatial summation this results in a shallow depression in acoustic output, compared to the in-wall installation where there is no interference.

If the loudspeaker were positioned at a distance from the wall, as when using a mounting bracket, the dip would begin at a lower frequency. The greater the separation, the lower the frequency at which the depression begins. The more directional the loudspeaker, the narrower the frequency range over which it exists. This is the mechanism of the classic adjacent-boundary effect.

What we need are loudspeakers designed with knowledge of how they are to be mounted, where they are to be placed. The idea of a universally applicable, one-type-does-all, loudspeaker is an obsolete concept, but it is the basis of most of today's designs: bookshelf loudspeakers that don't work in bookshelves, free-standing loudspeakers that for practical reasons cannot avoid at least some adjacent-boundary problems. It would seem that there is an opportunity for something different.

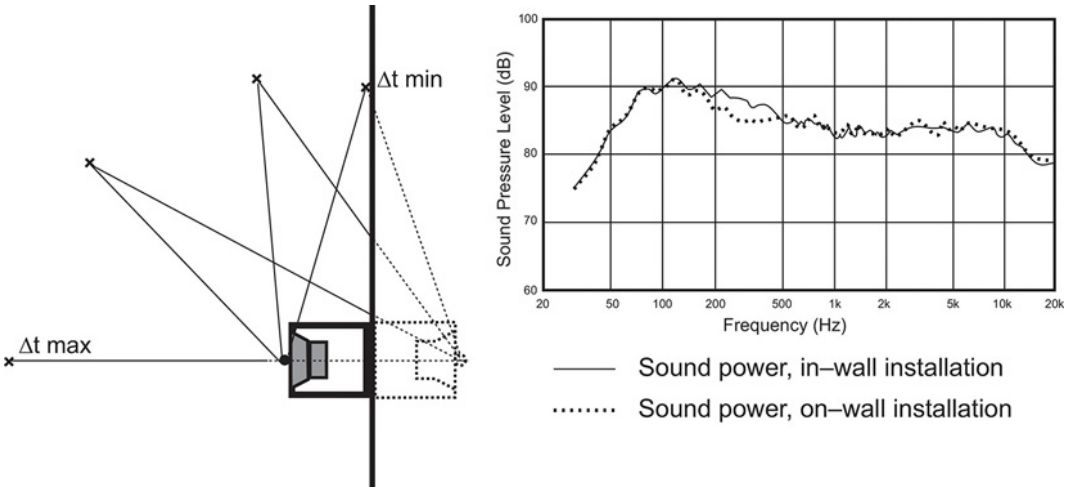


FIGURE 13.10 *A direct comparison of sound power measurements made with the same loudspeaker mounted in and on a wall, as shown in Figure 13.8. The ray diagram applies to the on-wall situation, showing the direct sound paths from the loudspeaker and the reflected sound paths from its acoustical mirror image behind the wall.* ⏎

13.5 BOUNDARY-FRIENDLY LOUDSPEAKER DESIGNS

Having expounded on the problems created by adjacent-boundary effects, it was no surprise that Allison (1974) proposed a loudspeaker design that minimized the effects. Figure 13.11 shows the configuration of the loudspeaker and how it was intended to be placed in a room. It also shows a later design influenced by the same concepts.

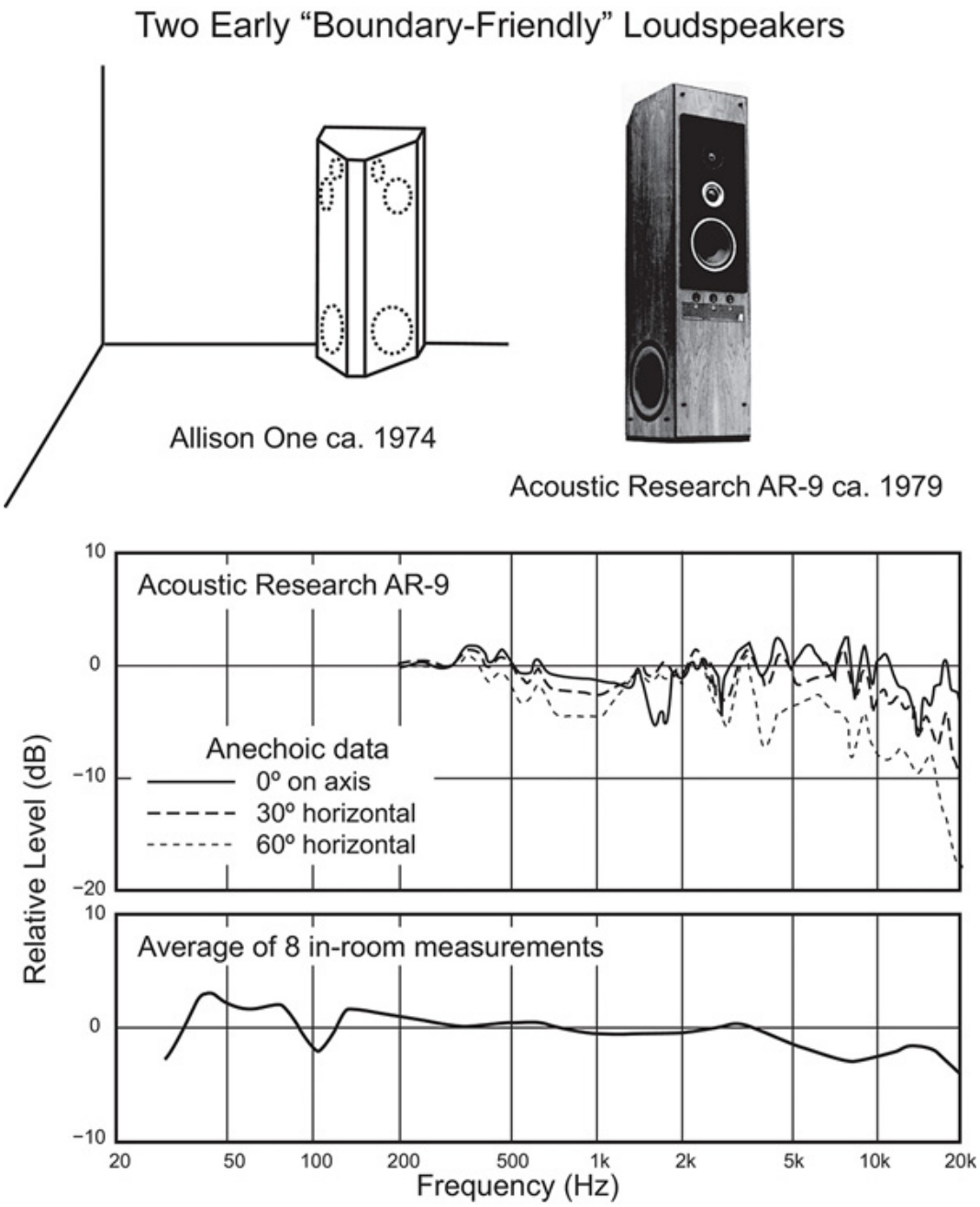


FIGURE 13.11 *Two early boundary-friendly loudspeaker designs. The Allison One loudspeaker as described in Allison (1974, figure 16) and the Acoustic Research AR-9. The anechoic curves are truncated at 200 Hz because of chamber errors in measurements of that period.* ⏎

The common factor in these designs is the placement of woofers close to both the floor and the wall behind the loudspeaker, incorporating those boundaries into the loudspeaker design. This maximizes the solid angle gain and minimizes adjacent boundary problems. The effect of the side wall is minimized by placing the loudspeaker some distance away from it, if possible.

In the Allison One the two sets of drivers at 90° to each other were intended to approximate a hemispherical radiation pattern. It was a very thoughtful design but, unfortunately, I have no measurement data on it. The duplicated midranges and tweeters probably resulted in a softening of stereo imaging, which would be compatible with the spaciousness that would accompany the wide dispersion. Personal preferences enter the scene.

The Acoustic Research AR-9 respected Allison's adjacent-boundary concerns with the woofer placement. A conventional forward-firing arrangement of drivers completed the system. Some anechoic data are shown, along with a spatially averaged room curve. It was a good-sounding loudspeaker.

Controlling adjacent boundary effects improves the uniformity of sound power radiated into a room, but it does not exercise control over the large effects of standing waves at the listening locations; these would be strongly room dependent. The room curve shown was the result of several combinations of loudspeaker and microphone locations, attempting to reveal the performance of the loudspeaker-and-adjacent-boundary combination. The result confirms that the design was successful, even with residual standing wave effects included.

Loudspeakers designed to function well when close to boundaries offer an advantage given the present popularity of on-wall flat-screen video displays. Low-profile loudspeakers situated on or close to walls are clearly desirable and function very well when designed for this mounting.

The surround loudspeaker shown in Figure 13.12(a) is an example of a class of products developed to cater to home theaters. It was designed with on-wall mounting in mind, and considerable thought went into the physical

layout and the crossover network to control the acoustical interactions of the drivers with each other and with the wall behind.

In spite of an appearance suggesting bidirectional radiation, over most of the frequency range this loudspeaker behaves as a hemispherical radiator. The measured curves in (b) are all very similar. For surround-channel loudspeakers mounted on side and rear walls this ensures delivery of similar direct sound to all parts of the audience. These bipole (bidirectional in-phase) designs are also appropriate for the current immersive systems; the dipole versions are not (see Section 8.5.3).

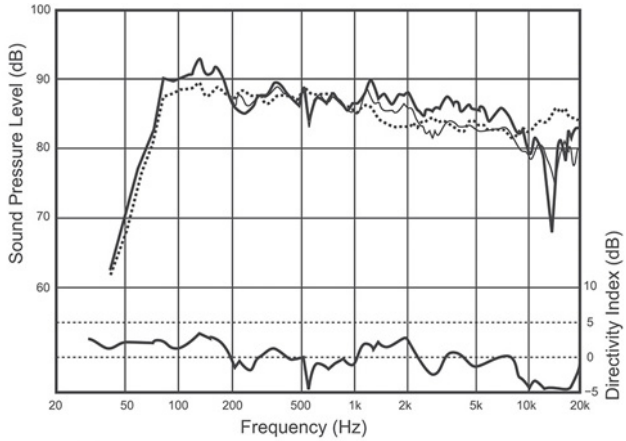
In-wall, flush-mounting is excellent and, with good design, on-wall configurations work very well. As shown in this example, they allow for nearly hemispherical radiation. Many surround loudspeakers are designed in this fashion, a welcome trend.

Ironically, it is the front loudspeakers, the most important ones, which routinely are designed with little or no regard for the adjacent-boundary settings into which they will be placed. It is a welcome trend to see in-wall loudspeakers that can compete with free-standing loudspeakers in sound quality and power output. They are widely used in custom home theaters. I used one as a center channel (the most important channel) in one of my home theaters. It was in a custom enclosure.



Infinity Beta ES250 ca. 2005

(a) A bidirectional in–phase (aka “bipole”) surround loudspeaker designed for on-wall mounting



(b) on axis
±30° horizontal ±10° vertical window
sound power

FIGURE 13.12 *(a) An Infinity Beta ES250 on-wall wide-dispersion surround loudspeaker with switchable directivity. (b) Frequency response measurements on axis and averaged over the listening window ±30° horizontal and ±10° vertical (nine curves) compared with the total sound power. The directivity index was computed using the listening window as the reference. For the measurements the loudspeaker was used in the "bipole" mode: both sets of drivers were radiating in phase with each other, and it was mounted on a large section of wall in an anechoic chamber.* ⏎

In control rooms, it has been common practice for decades to mount the main front monitor loudspeakers in some form of half-space mounting. Eargle (1973) and Mäkivirta and Anet (2001) are examples of advocates of this form of installation. Some rooms, because of the location of the viewing window into the studio, force the loudspeakers to be installed in an overhanging structure placed against the ceiling and with the loudspeaker surface some distance forward of the window and remainder of the wall. In the audio industry this is called soffit mounting, although the word normally has a different meaning. It is a corruption of half-space mounting; consequently, non-optimal boundary interactions may be anticipated.

Section 12.2.4 discusses a variety of vertical arrays and their interaction with floors—another adjacent-boundary effect that physics can help to manage.

13.6 LISTENERS ALSO HAVE BOUNDARIES

A lot of effort and talk goes into getting the right acoustical setting for the loudspeakers, but little, it seems, for the listener. The room boundaries around the listener also matter. In perusing a high-end audio magazine, I saw a photo of a reviewer's listening room. My first impression was that it was probably not a bad room, the loudspeakers were well away from the boundaries, the walls were geometrically irregular with many large shelving

units full of books and records, and there was what looked like a good carpet. He probably had developed "reference" locations for loudspeakers he was reviewing, but his "reference" listening position had not benefited from the same amount of attention. The solitary chair was within a few inches of the back wall. Everything that reviewer heard and commented on was corrupted by rear wall reflections. Pity.

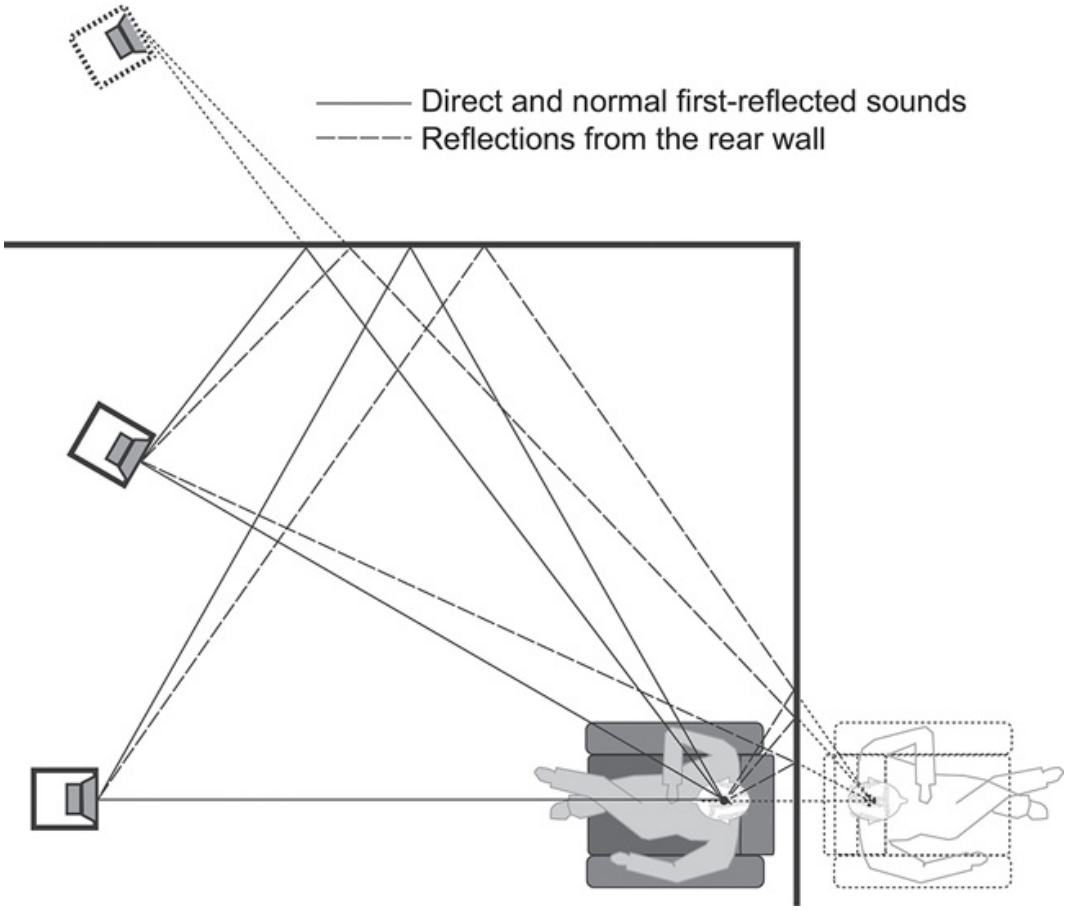


FIGURE 13.13 *A ray-tracing diagram showing the additional reflections caused by proximity to a reflecting rear wall. Not only loudspeakers have acoustical mirror images.* ⏎

It is an unfortunate fact of many home listening rooms, and I have experienced it in some hotel-room audio-show demonstrations. The ideal solution is to move the listener away from the wall, but with limited space this may not be possible. Figure 13.13 shows what happens; a whole new

set of reflections is created, and the ones immediately following the direct sounds from the loudspeakers are very early and very strong. There will be ceiling reflections as well.

In Figure 13.13, multiple reflections from the rear add poorly correlated sounds to those arriving from the front. In addition to the timbral colorations added by the acoustical interference, there will be degraded image and soundstage quality.

When confronted with such a situation, a persuasive demonstration is to have the listener play a stereo recording, asking him or her to focus on imaging and soundstage. Then hold a thick fabric-upholstered cushion or pillow filled with fibrous stuffing behind the head. It has been my experience that most listeners notice an improved clarity of sound and more precise imaging. Obviously, the permanent solution to this problem is to attenuate the rear-wall reflections. Place a panel of thick—up to 6 inches (150 mm) if possible— fibrous absorbing material covering a significant area behind, beside, and above the listener's head. I have seen this hidden behind a visually attractive, acoustically transparent, wall hanging. It won't be perfect, but it will be an improvement.

REFERENCES

Allison, R. F. (1974). "The Influence of Room Boundaries on Loudspeaker Power Output", *J. Audio Eng. Soc*., vol. 22, pp. 314–320. ⏎

Allison, R. F. (1975). "The Sound Field in Home Listening Rooms, II", *J. Audio Eng. Soc*., vol. 24, pp. 14–19. ⏎

Allison, R. F. and Berkovitz, R. (1972). "The Sound Field in Home Listening Rooms", *J. Audio Eng. Soc.*, vol. 20, pp. 459–469. ⏎

Angus, J. (2010). "Closed and Vented Loudspeaker Alignments That Compensate For Nearby Room Boundaries", *Audio Eng. Soc. 128th Convention*, Preprint 7990. ⏎

Ballagh, K. O. (1983). "Optimum Loudspeaker Placement Near Reflecting Planes", *J. Audio Eng. Soc*., vol. 31, pp. 931–935. See also letters to the editor, vol. 32, p. 677 (1984). ⏎

Eargle, J. (1973). “Equalizing the Monitoring Environment”, *J. Audio Eng. Soc*., vol. 21, pp. 103–107. ⏎

Mäkivirta, A. V. and Anet, C. (2001). “A Survey Study of In-Situ Stereo and Multi-channel Monitoring Conditions”, *111th Convention, Audio Eng. Soc*., Preprint 5496. ⏎

Pedersen, J. A. (2003). “Adjusting a Loudspeaker to its Acoustic Environment—the ABC System”, *115th Convention, Audio Eng. Soc*., Preprint 5880. ⏎

Waterhouse, R. V. (1958). “Output of a Sound Source in a Reflecting Chamber and Other Reflecting Environments”, *J. Acoust. Soc. Am*., 30, pp. 4–13. ⏎

CHAPTE
R 14

Below the Transition Frequency: Optimizing the Experience

Todd Welti and Floyd Toole

DOI: 10.4324/9781003477495-15

"Small rooms" are notorious for bad bass. Sustained bass notes fluctuate in level as they change frequency. Bass transients "boom" and lack the tightness that can be experienced in live performances. All these audible characteristics change as one moves around the room or sits in different seats. These problems exist in recording control rooms used in the creation of music, movie and television sound, and the multitudes of domestic spaces devoted to stereo listening and home theaters. In assessing the factors contributing to subjective judgments of sound quality, discussed in Chapter 11, it was shown that about 30% of the overall rating is contributed by factors related to low-frequency performance (Olive, 2004a, 2004b). It is a major problem in all of audio.

Our idealized objective is to achieve high subjective ratings for loudspeakers, and to do it consistently in different seats and different rooms. Up to a point it happens now, as discussed in Sections 5.4 and 5.5, but even so there was evidence of variations in judgments due to irregularities in bass performance. Humans can adapt to a certain amount of bass misbehavior, but there are limits.

A strategy is needed that can ensure the delivery of similarly good bass to all listeners in all rooms. As discussed in Section 1.7, the "circle of confusion" argues that achieving such consistency is a necessary objective for the entire audio industry. Ideally, we want recording professionals and consumers to hear the same quantity and quality of bass. Let us see how far this idea can be taken.

14.1 OBJECTIVES AND ASSUMPTIONS

There are three primary objectives in optimizing low-frequency reproduction in rooms *with more than one seat.* In order of importance, they are:

1. Reduce seat-to-seat variance of frequency response.
2. Maximize low-frequency efficiency.
3. Achieve smooth frequency response.

The first one is the most difficult to achieve and is the primary focus of this chapter.

Throughout this chapter, certain nominal assumptions are made for reasons which will become clear as we go along:

1. The emphasis is on rectangular rooms (see Section 14.4.2 for a short discussion of non-cuboid rooms).
2. Multiple low-frequency sources are assumed to be identical and driven coherently, not in stereo left/right or other multichannel mode. We are

assuming that bass management is used.

3. The focus is on low-order axial modes, though the principles apply to other types of modes.
4. It is assumed that proper low-frequency "global" equalization is to be applied after modal optimization. See Section 14.5.9 for discussion of system equalization at low frequencies.
5. We are focusing on magnitude response (room modes are generally minimum phase anyways).

14.2 THE BASICS OF ROOM RESONANCES AND STANDING WAVES

All the bass we hear, no matter how excellent the woofers/subwoofers may be, is filtered through room resonances. They affect what we hear at home and in our cars, adding to the problem of communicating the recorded art as it was intended. Discussions of the effectiveness of "room equalization" or "room calibration" are confounded by the fact that commercial algorithms differ widely in the resolution of the measurements they do, the resolution of the equalization employed, and in the algorithmic rules by which the corrective filters are created, and last but not least, the target performance being strived for. This topic was discussed in Section 12.3.3. In order to grasp and utilize the low-frequency optimization methods discussed here, a fundamental understanding of standing waves is required. There is a basic explanation in Section 9.9.2.

Figure 14.1 shows for a rectangular room (a) the orientations of the three axial modes, each of which are involves only one pair of parallel surfaces; (b) the three tangential modes, each of which involves two pairs of surfaces; and (c) one of many possible oblique modes which involve all surfaces. Because some energy is lost at every boundary interaction, modes that complete their "cycle" of the room with the fewest reflections are the most energetic and decay the slowest. The axial modes are therefore the most energetic and problematic, followed by the tangential modes and the

oblique modes. It is rare for an oblique mode to be identifiable as a problem in a room. Tangential modes can be found in rooms having significantly reflective boundaries, or when multiple sources are appropriately located. Axial modes are omnipresent, and they are the usual culprits in bass problems in small rooms. Fortunately, they are easily calculated, as described in the box.

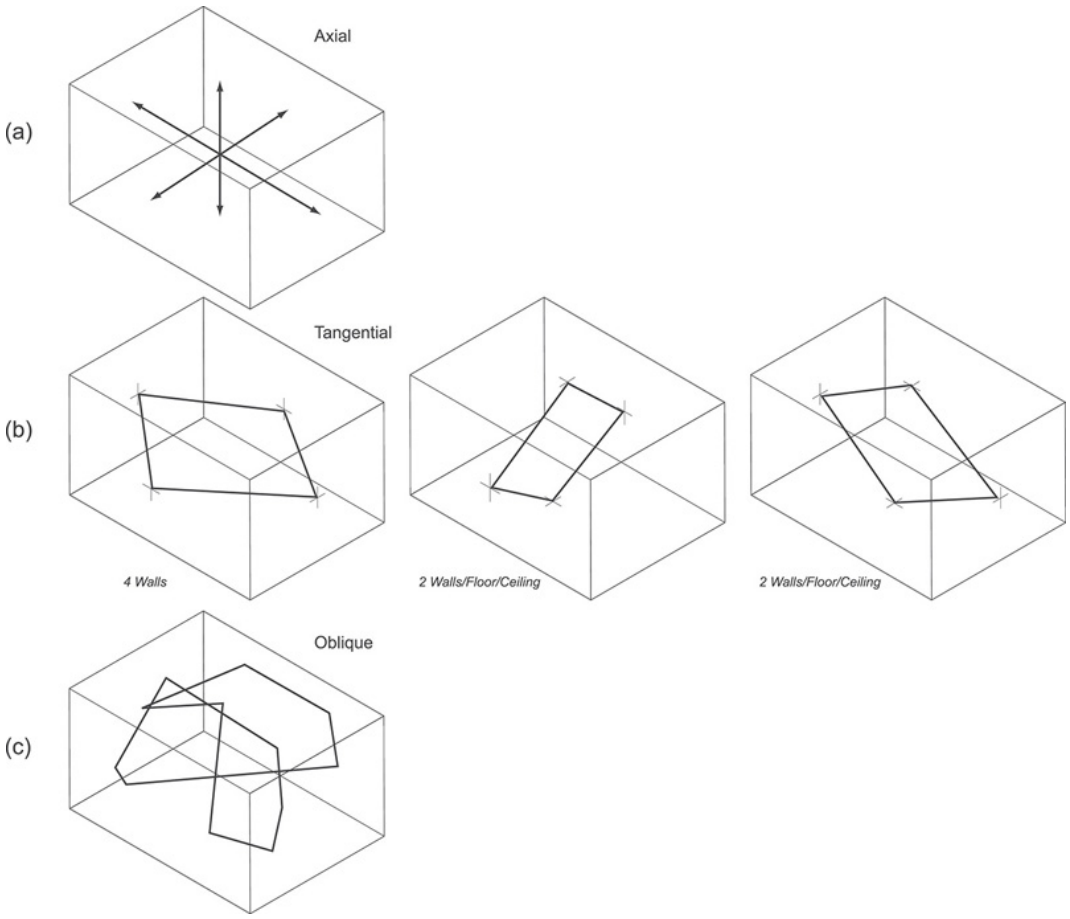


FIGURE 14.1 *The three classes of modes shown in a cuboid room: (a) axial: length, width and height; (b) tangential, each of the three involving two pairs of parallel boundaries and ignoring the third pair. In (c) is shown one of the many possibilities for oblique modes that involve all surfaces.* ⏎

"Back of the envelope" acoustical calculations. To compute the frequencies at which axial standing waves occur, simply measure the distance between the walls (this is one-half wavelength of the lowest resonance frequency), multiply it by two (to get the wavelength) and divide the speed of sound by that number in whatever units the measurements were done (1131 ft/s [345 m/s]). The result is the frequency of the first-order mode along that dimension. All higher-order modes are the result of multiplying this frequency by 2, 3, 4, 5, etc. For example: a room is 22 feet long. The first-order length resonance occurs at 1131/44 = 25.7 Hz. Higher-order resonances exist at 51.4 Hz, 77.1 Hz, 102.8 Hz and so on. Repeat this for the width and height of the room, and all of the axial modes will have been calculated.

One should note that the calculated and actual mode frequencies in a real room may differ. Still, it is useful to know where they *should* be if one is trying to track down the origin of an irritating bass boom and from that to decide on a remedy (see Figure 5.2). Figure 14.2 shows the equation giving the modal frequencies for a rectangular room of known dimensions.

$$f_{n_x n_y n_z} = \frac{c}{2}\sqrt{\left(\frac{n_x}{l_x}\right)^2 + \left(\frac{n_y}{l_y}\right)^2 + \left(\frac{n_z}{l_z}\right)^2} \qquad \text{[Equation 14.1]}$$

$f_{n_x n_y n_z}$ = the frequency of the mode defined by the integers applied to dimensions x, y and z (length, width and height).

n_x, n_y, n_z = integers from 0 to ∞ applied to each dimension: x, y, z.

l = dimension of the room in ft (m)
c = speed of sound: 1131 ft/s (345 m/s)

Examples of mode identification:

f1,0,0 is the first–order length mode (x dimension)
f0,2,0 is the second–order width mode (y dimension)
f0,0,4 is the fourth–order height mode (z dimension)
f1,2,0 is a tangential mode (involving two dimensions)
f1,3,2 is an oblique mode (involving all three dimensions)

FIGURE 14.2 *The general Equation 14.1 for calculating all possible modes in a rectangular room.* ⏎

Figure 14.3 shows all of the expected modes for a small rectangular room. This is the kind of information yielded by any of several computer programs that run the preceding equation. Not all of these modes will be problematic in practical circumstances, but they exist, and knowing the frequencies at which they occur can be helpful in analyzing specific installations.

The lowest mode will be associated with the longest room dimension. Another general feature of rooms is that modes are generally sparser and more isolated at low frequency. At higher frequencies, modes are more numerous and overlap to the point of not being distinguishable. See Section 5.1 for more perspective.

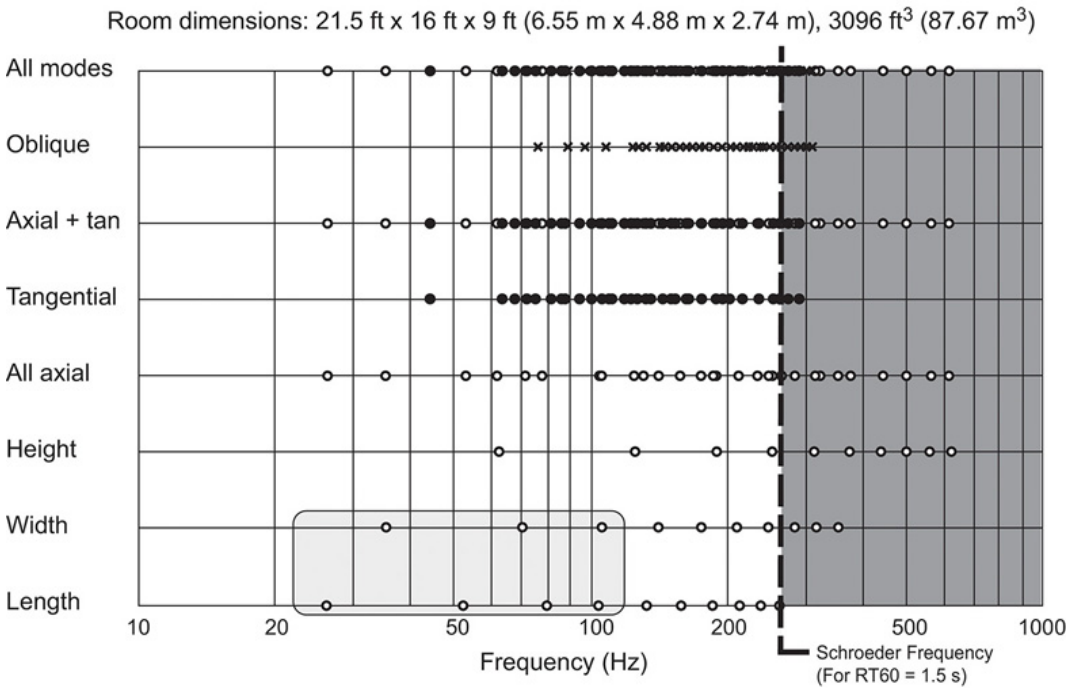


FIGURE 14.3 *The low-order modes for a small rectangular room. This is based on the presentation of data by a Microsoft Excel program which is available for download at www.harman.com/audio-innovations and in OR Chapter 14. The darker shaded area is above the Schroeder frequency assuming RT60 = 1.5 s, so all of the modes have not been calculated. The light shaded area indicates the modes that are of particular interest since they are isolated axial modes. Note that height modes are less critical since they do not cause seat-to-seat variation (assuming a flat floor and ceiling).* ⏎

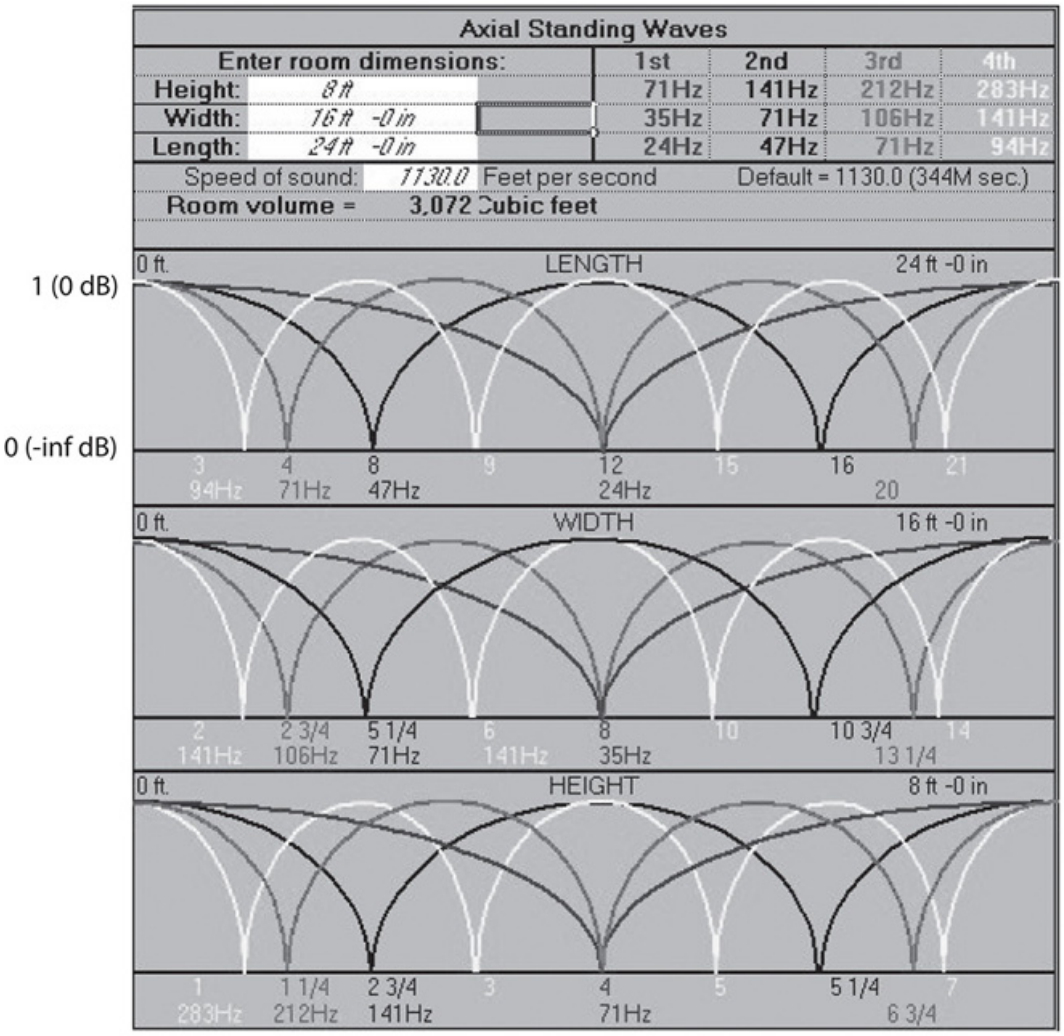


FIGURE 14.4 *Example of standing wave patterns along the length, width and height for a rectangular room. Vertical axis for each plot goes from 0 at bottom to 1 at top (or −□ to 0 dB if you prefer). From the same mode calculator referenced in Figure 14.3.* ⏎

Figure 14.4 is the output diagram of the Harman calculator showing the standing wave patterns for four orders of axial modes. Both sets of data can be useful in identifying problem room modes. Standing wave patterns for tangential and oblique modes are harder to visualize but otherwise behave the same. See OR Chapter 14 for more details.

14.3 A MORE COMPLETE UNDERSTANDING

We have shown how to calculate expected modal frequencies and how axial modes vary along a single dimension. This does not give a complete picture. In this section we delve a little deeper, with the expectation that this will help better understand *why* the various optimization methods we discuss work. This discussion includes some generalizations and simplifications. A more complete picture can be found in, for example, Kleiner and Tichy (2018).

Room modes do not operate only as a function of space but simultaneously as a function of frequency. There exists an equation for rectangular rooms which gives the effect over frequency and over space for *each* separate mode and then you can sum them up to give expected results for any given room dimensions and source/receiver positions. The full equation can be found for example in Walker (1992) and was used in many of the results shown later in this chapter. It is a bit complicated, and several different forms are commonly found in the literature.

The equation can be stated in a much *simplified* form as in Figure 14.5 Equation 14.2, which can aid in a fuller understanding. For each individual mode, assuming source and receiver locations are fixed at positions S and R the response $p_n\ (f)$ (magnitude) can be conceptualized as follows. Note that in the full form the equation gives actual sound pressure level as output, but we are simplifying considerably and so the output is considered in a relative sense. Also, for simplicity we are assuming an ideal source with no roll-off over frequency.

Here $p_n\ (f)$ is the modal response for a single mode n in one dimension. The factors Ψ_S and Ψ_R are coupling factors that describe how well the source (S) and receiver (R) couple into or out of each room mode based on their positions within the room, and f_n is the modal frequency for the particular mode being considered. So, Ψ_S and Ψ_R are the same functions we have already seen in Figure 14.4 (for the more mathematically inclined these are based on cosine functions). The D in the denominator is a small damping factor that prevents the denominator from going to zero (implying

infinite results) when $f = f_n$, i.e. at the mode resonant frequency. What this conceptualized equation shows is that the response of a single mode can be separated neatly into a spatial component in the numerator, and a frequency component in the denominator.

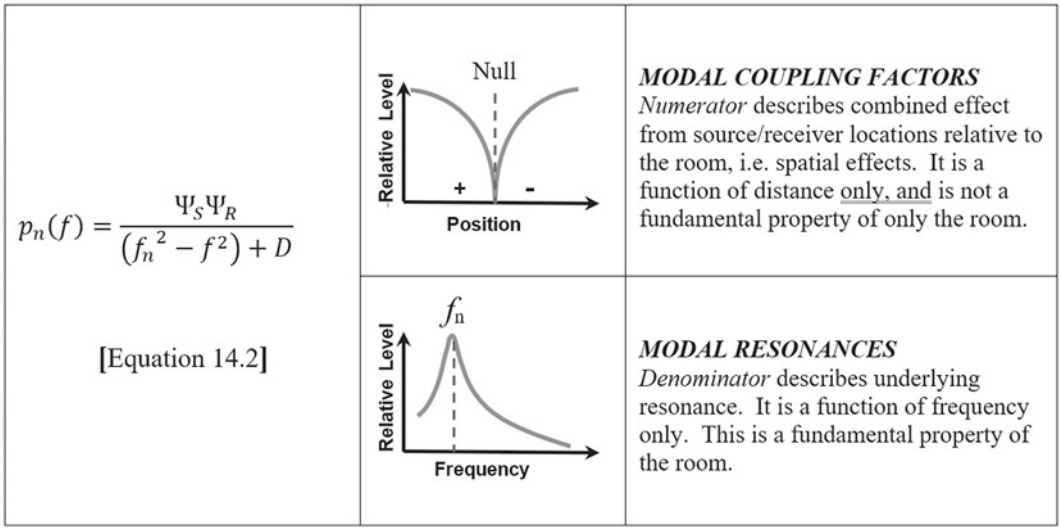


FIGURE 14.5 *A highly conceptualized version of the full analytical equation for a single room mode in a rectangular enclosure, highlighting the spatial effects of the numerator (including polarity inversion across the null) and the frequency effects of the underlying modal resonance in the denominator. This is for a single mode n in one dimension.* ⏎

14.3.1 Modal Coupling Factors

Let's look at Figure 14.6, which shows two examples of first and second axial modes. A source and receiver are shown in each case. It is important to note that the loudspeaker source is a pressure source, as in a closed box or reflex design, not a velocity source, as in a dipole. Human listeners are pressure receptors.

We can visualize the coupling into and out of each mode depending on where the source and receiver are relative to the room dimension. This is represented by the numerator of Equation 14.2 in Figure 14.5. Note that as you move across any null in the magnitude response, the polarity is

inverted. This is a fundamental property of resonances in general, and room modes are no exception. In the first example, the subwoofer source couples into both modes fully, and the receiver located at center of the room does not couple into the first mode at all, but couples fully into the second mode. In the second example the source couples moderately into both modes, and the receiver couples moderately into the first mode and not at all into the second. The results of these mode couplings are summarized in Table 14.1, and the combined coupling Ψ_S and Ψ_r are shown. These are just examples. Hint: "X" marks the spot! Note that the vertical scale is in absolute terms on the left, and equivalent dB values on the right.

$$p_n(f) = \frac{\Psi_S \Psi_R}{(f_n{}^2 - f^2) + D} \qquad \text{[Equation 14.2]}$$

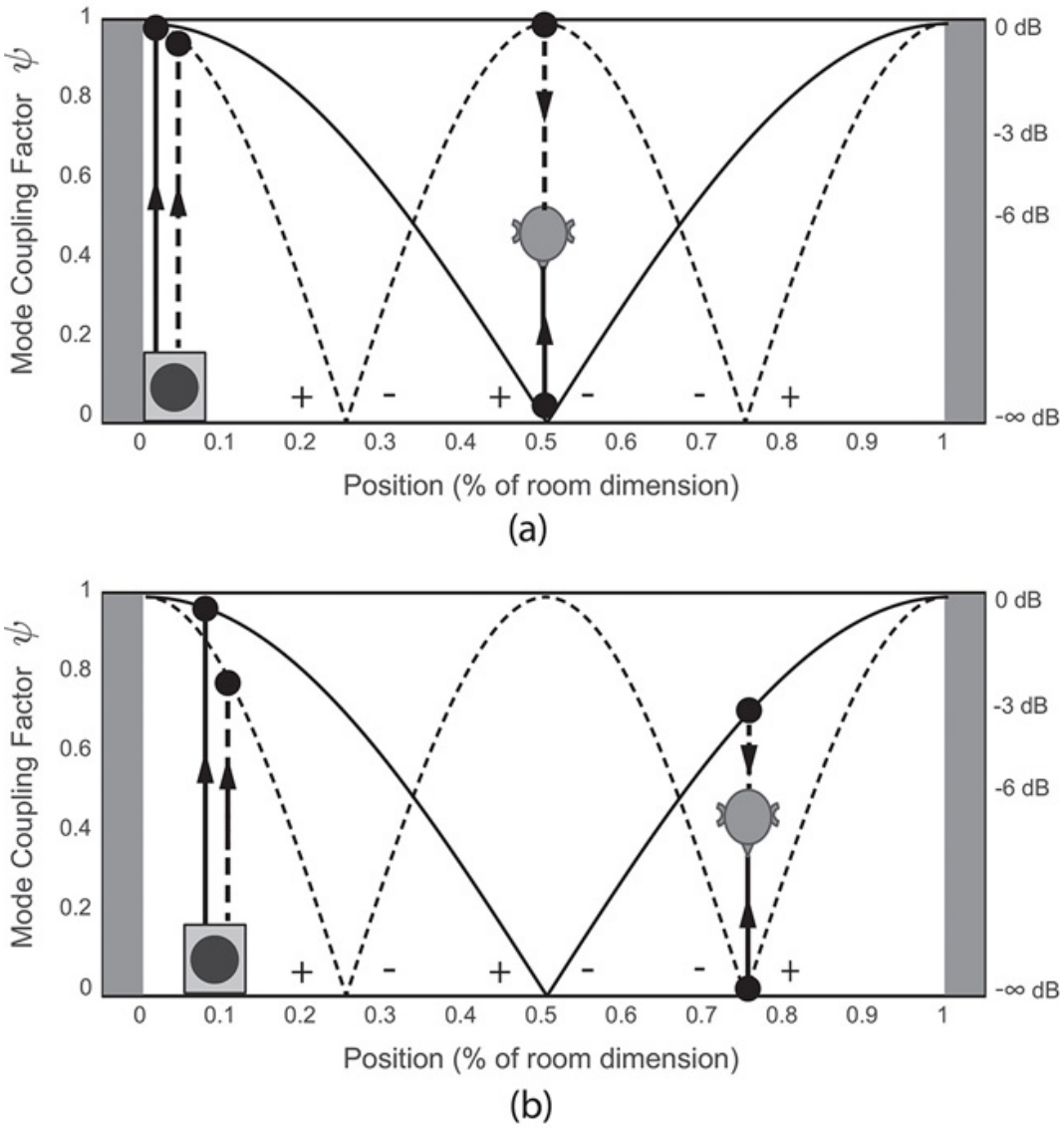


FIGURE 14.6 *Two examples, (a) and (b), of standing wave patterns for first two axial modes between two walls in a room. The standing wave pattern is also the coupling factor for source and receiver: Ψ_S and Ψ_R. First and second axial modes n = 1 and n = 2 are shown. Arrows indicate coupling into or out of the modes. Polarity inversions on either side of nulls are also shown.*

Question: Where would you put the receiver in Figure 14.6 (a) to hear both modes equally?

To summarize, the net effect of the mode at or near the modal frequency depends on the positions of both source and receiver, and each mode is different. These spatial effects are independent of frequency, i.e. they occur at *all* frequencies for a given mode. This will become clearer as we go along.

TABLE 14.1 Summary of modal coupling factor examples illustrated in Figure 14.6. ⏎

Example (a):

Mode [n]	$y(S)$	$y(R)$	$y(S)\,y(R)$	$y(S)\,y(R)$ [dB]
1	1	0	0	$-\infty$
2	1	−1	−1	0 @ 180 deg.

Example (b):

Mode [n]	$y(S)$	$y(R)$	$y(S)\,y(R)$	$y(S)\,y(R)$ [dB]
1	.95	−.7	−.67	−3.5 @ 180 deg.
2	.75	0	0	$-\infty$

14.3.2 Modal Resonances

Now let us look at the denominator of Equation 14.2, which represents the modal resonance in frequency. For any given mode, as the frequency of a source approaches the modal frequency f_n, the denominator gets smaller and thus the output reaches a peak. This will look something like what is shown

in Figure 14.5. This resonance in frequency applies equally for all source and receiver positions. This is the beauty of our equation, it allows us to separately consider the spatial and frequency effects of real modes in real rooms.

Keep in mind that there is a "Zeroth" order mode 0,0,0, commonly known as room gain or, more accurately, room pressurization mode that contributes to the modal sum. This represents a first-order compliance effect caused by an enclosed space, occurring below the first axial mode. In our graphs, we illustrate the room pressurization mode for a theoretical sealed room. However, since most rooms are not sealed, this effect will be less pronounced, and the response would actually go to zero at the lowest frequencies.

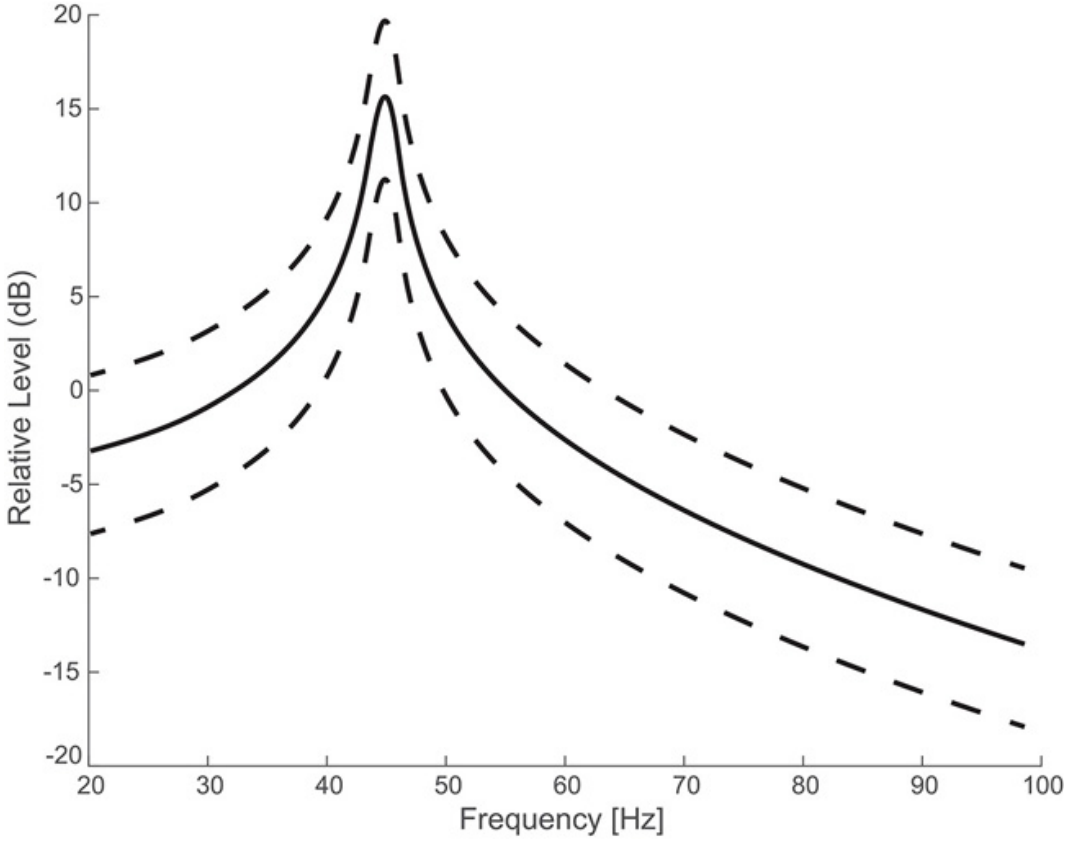


FIGURE 14.7 *Effect of total modal coupling factor ($\Psi_S\Psi_R$) on a single modal resonance, shifting it up or down.*

Now let us look at the combined Equation 14.2 for one single mode. Figure 14.7 summarizes the effects of modal coupling on one modal resonance. In a real room the mode shapes are fixed, and simply shift up and down as you move source or receiver around to change the modal coupling factors. The effect can be complex as discussed in the next section.

14.3.3 Simplifying a Complex Situation

So far we have been looking at only one or two axial length modes. If only real rooms were that simple! Equation 14.2 is only a building block, and to be able to represent a real room we must sum up all the possible modes. This is shown in general terms in Equation 14.3, which says that the total modal response in a room is the result of summing all modes—all types of modes from all surfaces and all orders of modes for *each* source. Even if we limit ourselves to only first axial modes, we still must sum up modes associated with each pair of walls and ceiling: 1,0,0, 0,1,0, and 0,0,1. These don't line up neatly on a one dimensional plot like our previous figures. If that were not enough, modes often overlap, and in complex ways due to differing amplitudes and phase effects. To fully account for all of this would likely require a computer simulation. How can we simplify things?

$$P(f)_T = \sum_{n_x n_y n_z} p_n(f) n_x n_y n_z = (0:N, 0:N, 0:N) \qquad \text{Equation 14.3}$$

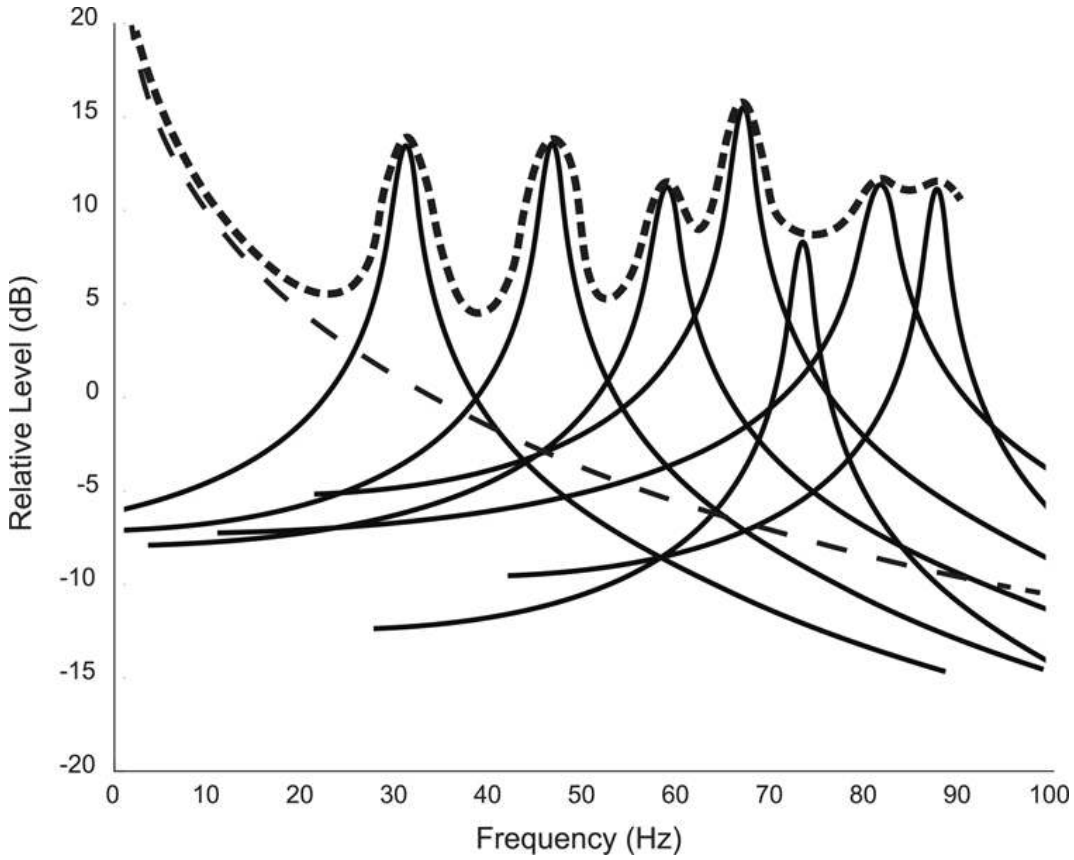


FIGURE 14.8 *Generalized summation of room modes representing average levels for receivers throughout the volume of a room, with subwoofer exciting all modes equally.*

One might be tempted to make a "skyline" plot tracing the tops of all the predicted modes in the frequency domain, such as shown in Figure 14.8. This type of plot can be found in many acoustics textbooks. It is valid only in the most general sense. If one were to put a subwoofer in one corner of a rectangular room such that all modes were fully excited, and measure throughout the entire volume of the room, one might see something resembling this.

A better way to simplify is to focus on modes which are most prominent and most isolated, which often are the ones which cause the worst problems, and are often the lowest order axial modes. Unlike higher frequency modes, where they may combine in complex ways, isolated lower modes are easier to analyze and manipulate to our advantage.

Figure 14.9 shows a conceptual example, where we focus on one low-order axial mode which has excessive energy and is well separated from nearby modes. Let's assume that this particular mode causes undesired seat-to-seat variation in the 45 Hz region, indicated by the upper and lower dashed lines in the region of the peak separated by about 10 dB. Equalization is not a solution! A single band-stop filter applied to the system at 45 Hz can do nothing other than reduce the output similarly at all seats, having no effect on seat-to-seat variation.

It can be seen that when the 45 Hz mode is lowered by reducing the source coupling factor, the neighboring modes and room pressurization mode will combine to "fill in" the now lower parts of the range between dotted lines, thus compressing the range of variation from 10 dB to around 5 dB. *The seat to seat variation is thus reduced.*

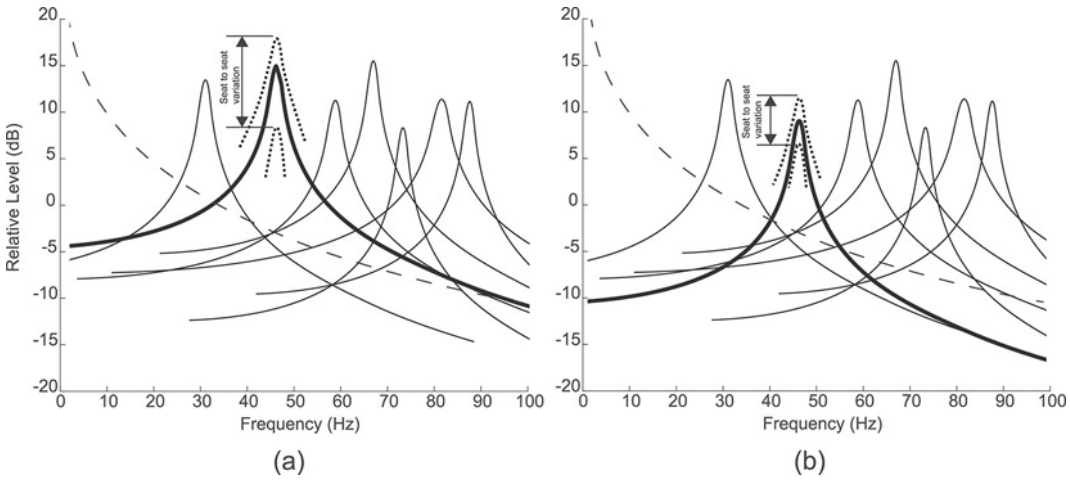


FIGURE 14.9 *Individual modes for an example room, with room pressurization mode resonance also shown (heavy dashed line). The 0,1,0 mode (heavy solid line) is a prominent and isolated mode which causes seat-to-seat variation (dotted lines). (a) shows unoptimized condition, and (b) shows result after selectively reducing energy in the 0,1,0 mode by, for example, moving the source (Ψs). The arrows show the seat to seat variation in this example for unoptimized (a) and optimized (b), near the resonances.*

What is needed is a way to reduce energy in the problem mode, shown in Figure 14.9(a), where there is a 10 dB range between the dotted lines. Let's assume for a moment we have achieved this. What is the result? In Figure 14.9(b) we see what happens when the mode in question is reduced in level. The response for that mode at all seats is reduced, but the neighboring (unaffected) modes and room pressurization mode will *tend* to fill in the response for seats with the lowest amplitude. This is what makes modal manipulation work. This is why it helps to be able to control each mode separately to the extent possible. This is the basic mechanism for reducing seat-to-seat variance. Sometimes it is as simple as in this example, or it may complicated enough that we must rely on generally optimum source configurations or other optimization methods.

14.3.4 Tools for Modal Manipulation—A Preview

What are our general tools for modal manipulation? Modal coupling factors (Equation 14.2 numerator) can be manipulated by moving sources and receivers within the room and/or by employing multiple sources at different positions with different coupling factors that combine. Frequencies of underlying modal resonances (Equation 14.2 denominator) may be manipulated by changing room dimensions. If source and receivers are kept in the same *relative* positions (X, Y and Z) during scaling such that coupling factors are not changed, all modal frequencies would simply be scaled inversely with the dimensions. In the extreme cases this can result in either increasing modal density such that they are no longer an issue, as in a large cinema, or moving problematic low-frequency modes up into higher registers where they are much *more* of an issue—a shower stall being an extreme example. If room dimensions are changed but source/receiver positions are fixed, then both modal frequencies and coupling factors would be changed. Finally, underlying modal resonances may be modified by changing the acoustical damping in the room, which changes the bandwidth, or Q, of the modes, and thus their overlap. This is loosely represented by D in Equation 14.2. Look to Kleiner and Tichy (2018) for

more details. These general methods are all discussed in more detail in following sections.

14.4 GENERALIZED OPTIMIZATION OF ROOM DIMENSIONS—DOES AN "IDEAL" ROOM EXIST?

A staple of low-frequency room acoustics lore is the idea that there is a set or sets of optimized room dimension ratios that will make rooms sound better. The general assumption is that having optimally spaced mode frequencies should minimize their detrimental impacts. It intuitively makes sense that multiple clustered modes might combine into a "supermode", or that a large gap in the mode structure could cause a hole in the combined response. People looking for a simplified path to success for room design might be tempted to simply adopt one of these "magic" room ratios, but as we shall see there is more to be considered. Before discussing where these ideas came from and whether they have merit, let us clarify what are the assumed conditions and criteria for a good room in this context. We need to ask exactly what it is we hope to "optimize".

Previous work going back decades has focused on even spacing of modal frequencies (defined in various ways) with benefits including flatness of response and avoiding closely spaced modes that can result in audible beating. The focus was on flat frequency response in early room ratio studies, like those by Bolt (1946), but the objectives at that time were not relevant to sound reproduction. It was assumed that if the room was optimized, then for any given source/receiver positions the result would be optimized in a general sense. These early efforts were concerned with optimizing reverberation chambers used for measuring power output of noisy mechanical devices. In audio, a comparable example is a recording studio, where the vocalists and instruments are at unpredictable locations and a well-blended sound is desired for microphone pickup.

The original research focused on the rooms themselves as separate entities, not considering the locations of the sound source (loudspeaker) or the receiver (listener), and as has been seen both are important in the

transfer of energy to resonances and to their audibility. In sound reproduction we normally *do* know where the loudspeaker is, and also the seats or seating area—for example monitoring rooms, home theaters, etc. with loudspeakers and seating at fixed/known positions. We also have access to equalization. These are two very different situations. As often happens, research in one domain is mistakenly assumed to be valid in another.

As mentioned at the start of this chapter, the focus here is seat-to-seat consistency. The question we will address is: *do optimized room ratios help improve seat-to-seat consistency?*

We are here considering only rectangular rooms. It could be argued that optimizing room mode spacing in non-rectangular rooms might have similar benefits as for rectangular (assuming there are any), but a simple set of ratios is obviously not practical for complex room shapes. In addition, it is often forgotten that there is a "range of validity" in terms of frequency and room volume. Bolt specified this explicitly, but this information is often omitted. This, and a general overview of past research can be found in Rindel (2021) and in OR Chapter 14.

14.4.1 Necessary Assumptions

For our current focus on sound reproduction, there is nothing to suggest that the acoustical performance of rooms can be generalized on the basis of their dimensional ratios, and that reliably hearing superiority of a "good" room is possible. The fact that highly respected researchers seem to have all arrived at different optimal ratios does not inspire confidence. The simple explanation is that there are problems with the basic assumptions underlying determinations of "optimum" room dimensions for domestic listening rooms or control rooms. The normal assumptions are:

- All of the calculated room modes are assumed to be simultaneously excited, and by a similar amount. This requires that the sound source be located at the intersection of three room boundaries, i.e. on the floor

or at the ceiling, in a corner. *Any* departure from this location will result in some modes being more strongly energized than others.

- Listeners can hear all of the modes equally. This requires that the head be located in another, diagonally opposite, three-boundary corner, *Any* departure from this listening position means that all of the modes will not be equally loud or even audible.
- The room is perfectly rectangular, with perfectly flat, highly reflecting (rigid and massive) walls, floor and ceiling.
- Flatness of response at one listening location is the primary metric for success. In fact, many researchers only look at modal spacing, not the actual flatness/smoothness of real responses.

These assumptions do not describe the circumstances or rooms in which we live and listen. In real rooms irregularities in geometry and variations in surface reflectivity affect the relative strength and frequencies of modes even if they had the potential of being uniformly energized. Source and receiver locations are major factors.

A relevant and eye-opening example from Todd's research (Welti, 2012) illustrates this. Most room dimension practitioners would agree that having rooms where one dimension is a multiple of another is bad. Louden stated: "It is well known that rooms with dimensions that bear simple ratios to one another have especially bad sound response" (Louden, 1971).

Having a square room would seem to be forbidden. However, using a custom room model, it has been shown that this is indeed true for one subwoofer in the corner but not true for certain optimized configurations such as four subwoofers at wall midpoints. Details can be found in OR Chapter 14. Multiple subwoofers change the rules.

There has been some further research in the last few years that has gone beyond modal spacing optimization. For example, in Meisner (2017) a numerical model was used (not generally available when Bolt and some other researchers did their seminal works). A large number of rooms of different dimensions were modeled, including many different source

positions throughout the room. These assumptions are more representative; however, the optimized metric is still a smoothness related one (not seat-to-seat variance), and receiver locations are assumed throughout the volume of the room, which is not directly applicable for our assumed conditions and goals.

In conclusion, it is not that the idea of optimum room ratios is wrong, it is simply that as originally conceived, it is irrelevant in our business of sound reproduction. See Section 14.5.6 for a discussion of a methodology to optimize room dimensions along with seating and subwoofer configurations together. This is more likely to be useful than employing "magic" room ratios.

14.4.2 Are Non-rectangular Rooms the Answer?

A recurring fantasy about rooms is that if one avoids parallel surfaces, room modes cannot exist. Sadly, it is incorrect. Among the few studies of this topic, Geddes (1982) provides some of the most useful insights. He found that "room shape has no significant effect on the spatial variations of the pressure response … the spatial standard deviations of the p^2 response is very nearly uniform for all the data cases [the five room shapes evaluated in the computer model]". Source location was a factor in the behavior of the modes of course, as was the distribution of absorption. "Distribution of absorption was far more important in the more symmetrical shapes—[a non-rectangular] shape did help to distribute the damping evenly among the modes" Geddes (2005). It seemed that the most effective modification to a rectangular room was to angle a single wall. Note also that so-called flutter echo is a different phenomenon which under certain conditions can benefit from angling one or more walls, although it is usually solved with absorbers and/or diffusers.

The behavior of sound in non-rectangular rooms is difficult to predict, requiring either scale models or powerful computer programs (e.g. NIRO, Redi Acoustics, https://rediacoustics.com/niro). Figure 14.10 shows two-dimensional estimated pressure distributions for modes in rectangular and non-rectangular spaces. Both shapes exhibit regions of high sound level,

and nodal lines where sound levels are very low. The real difference is that in rectangular rooms the patterns can be predicted using simple calculations. Consequently, most acousticians favor the relative simplicity of rectangular shapes.

14.5 SOLUTIONS FOR THE REAL WORLD

While earlier sections were a bit technical, here we will cover real-world issues and solutions. What are the actual deleterious *audible* effects of room modes in sound reproduction?

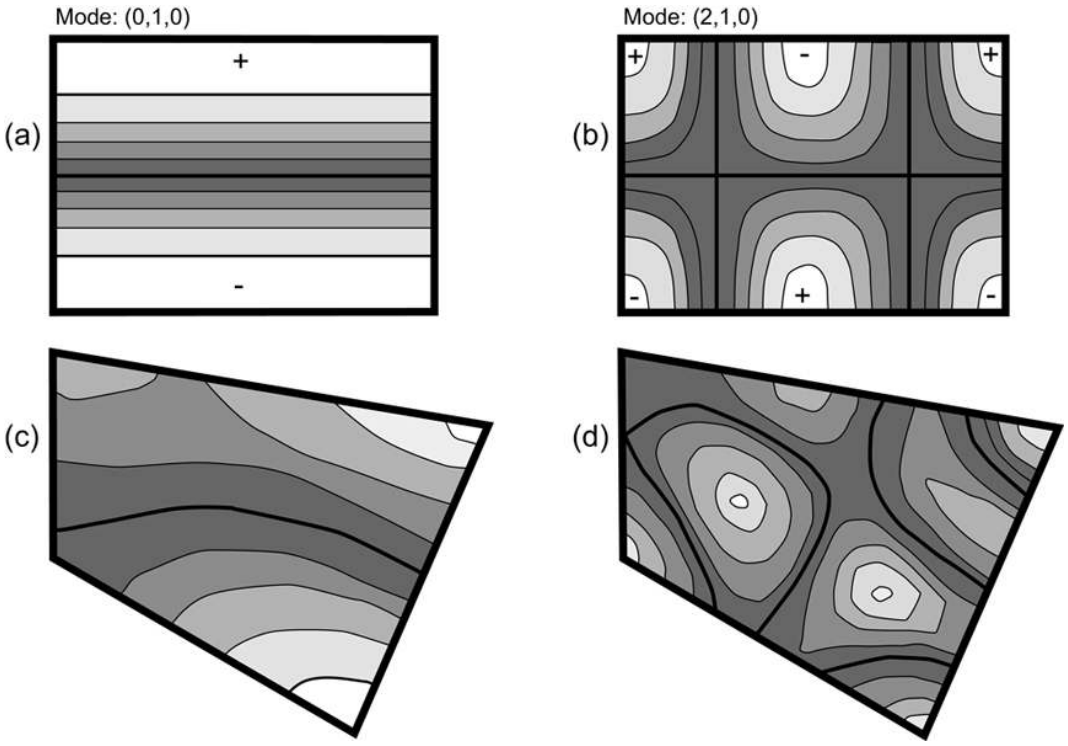


FIGURE 14.10 *Computed pressure distributions in rectangular and non-rectangular rooms, showing for each a simple mode and a more complex modal pattern. Pressure minima, nulls, are shown by heavy lines. Lighter shades indicate increasing sound pressure levels. Note that the instantaneous polarity of the pressure reverses at each nodal line (pressure minimum).*

There are three main categories of audible problems related to modes:

1. *Time-domain effects*: The prolongation of brief bass sounds, like a kick drum, by ringing that can occur at one or more resonance frequencies (bass "booms").
2. *Frequency-domain effects*: The amplification or attenuation of specific sustained bass notes (e.g. bass guitar, synthesizer, organ) at different frequencies. Pitch shifting is another subtle possibility as energy in a decaying mode at one frequency can migrate into a mode at an adjacent frequency.
3. *Spatial domain effects*: The variations in (1) and (2) that occur at different seats in the room. No two listeners hear precisely the same bass.

Occasionally there can be beating when adjacent modes trade energy back and forth in a chaotic manner over the course of their ringing.

As discussed above, the audible degradation depends on the combination of room dimensions, source and receiver positions, and the spacing and damping of the underlying room modes. Ultimately the effects also depend to some degree on the program material (see further discussion in Section 14.5.11). The result can be quite audible variations in loudness at different frequencies and locations in the room. At very low frequencies loudness changes by a factor of 2 with a mere 5 dB sound level change compared to about 10 dB at middle and high frequencies (see Section 9.4). Therefore, small changes in bass level seriously influence impressions of spectral balance; a deficiency of bass is often perceived as an excess of treble, and vice versa.

For decades the only way to improve bass sound in small rooms has been to employ low-frequency absorbers, a.k.a. bass traps, to damp the resonances. These are necessarily large and alter the visual décor. While this is an issue in domestic living spaces it may be more tolerable in professional recording studios or control rooms, or even in dedicated stereo listening rooms or home theaters. However, high-quality reproduced sound is also desired in rooms that serve as normal living spaces, where visual

appearances, even style, are a high priority. Consequently, as will be seen, the authors have been motivated to find alternatives to the traditional remedies.

The good news is that there are choices. The choice of method depends on your specific situation, requiring a balance between appearance, performance, and cost, while also considering which factors are available to be adjusted, such as room dimensions, speaker placements, or seating positions. Here are some of the popular options to addressing modal issues:

Passive Modal Resonance Modification

- Deliver energy to modes and dissipate some energy with low-frequency absorbers/bass traps.

Passive Modal Coupling Factor Modification

- Deliver energy to the modes and reduce the coupling of bothersome modes to the listener by optimizing the listening locations.
- Reduce the energy delivered to bothersome modes by optimizing the subwoofer location.
- Reduce the energy delivered to bothersome modes by using combined (often opposed) coupling factors of multiple subwoofers.
- Use well-known optimized subwoofer configurations for generalized multiple subwoofer mode manipulation.

Combined Passive Methods

- Optimize the *combination of* room dimensions, source and receiver positions to minimize bothersome modes.

Active Methods

- Manipulate modes using multiple subwoofers and computer optimized signal processing—rectangular rooms.
- Manipulate modes using multiple subwoofers and computer optimized signal processing—*all* rooms.

Equalization

- Global equalization at low frequencies after modal manipulation to minimize bothersome resonances and seat-to-seat variations, or equalization for a single listener with system as it is.

14.5.1 Deliver Energy to the Modes and Dissipate Some of That Energy With Absorbers

Here we are directly modifying modal resonances (Equation 14.2 denominator) as opposed to modal coupling factors. One way of thinking about it is that adding acoustical damping will in general make the individual modal peaks wider and smoother, i.e. lower the Q, as shown conceptually in Figure 14.11(a) and (b). This results in a flatter response on average and generally less seat-to-seat variance. Figure 14.11(c) shows an example in a rectangular room model with six seats, where the absorption coefficient on all surfaces was set to $\alpha = 0.1$ and then in (d) at $\alpha = 0.3$, again showing the effect of adding absorption. This roughly corresponds to absorption of double layer sheetrock and single layer sheetrock, respectively. Seat-to-seat variation at the lowest mode was reduced from 20 to 14 dB.

Looking at the time domain, added absorption also tends to damp modal decay, as shown in the example in Figure 14.12, taken in the venerable "Infinity Room" at Harman, in which the authors have done much research. In (a) all absorption has been removed from the room (except carpet). In (b) the normal complement of twelve 4 × 4 ft (1.2 × 1.2 m) 4 inch (100 mm)

thick fiberglass absorbers and four large tube traps are in the room. The effect is mostly as expected, with decays shortened. However, one can also see hints of more complex behavior, modes exchanging energy, etc. The devices used clearly were less effective in absorbing the lowest frequencies.

Adding absorption is a time-honored method, and it works. It is virtually a requirement in recording studios—spaces that become part of the microphone-captured "original" sound and must be relatively neutral. It is often used in control rooms and in dedicated home listening rooms but can be an aesthetic challenge in normal living spaces. To be maximally effective "bass traps" must be located in the high-pressure regions of the standing waves of bothersome modes meaning that they may be difficult to hide or disguise. The energy that is "trapped" is lost, of course; the total quantity of bass energy is reduced as a sacrifice to getting better quality bass. When one feels a wall, floor or membrane absorber vibrating, the mechanical energy in the movement has been removed from the sound field energy. Addressing specific bothersome modes without attenuating benevolent bass frequencies requires membrane absorbers tuned to those frequencies and/or absorbers placed in specific locations. Some thought is required.

In normal wood- or metal-frame buildings, i.e. many homes, the single layer of drywall is a moderately effective absorber at low frequencies. The existing room boundaries "trap" some bass energy. Figure 14.13 shows measurements of sound level vs. distance along the length of a listening room, at frequencies representing the first- and second-order axial modes along the length of the room. The room was perfectly rectangular, constructed of 2 × 6 inch (50 × 150 mm) studs with two layers of 5/8 inch (15.9 mm) gypsum board on the interior surface—it was designed with sound isolation as a consideration. Structurally, these walls are more massive and stiffer than those found in typical North American homes. The only opening was a heavy solid-core door located in the middle of the end wall on the right of Figure 14.13.

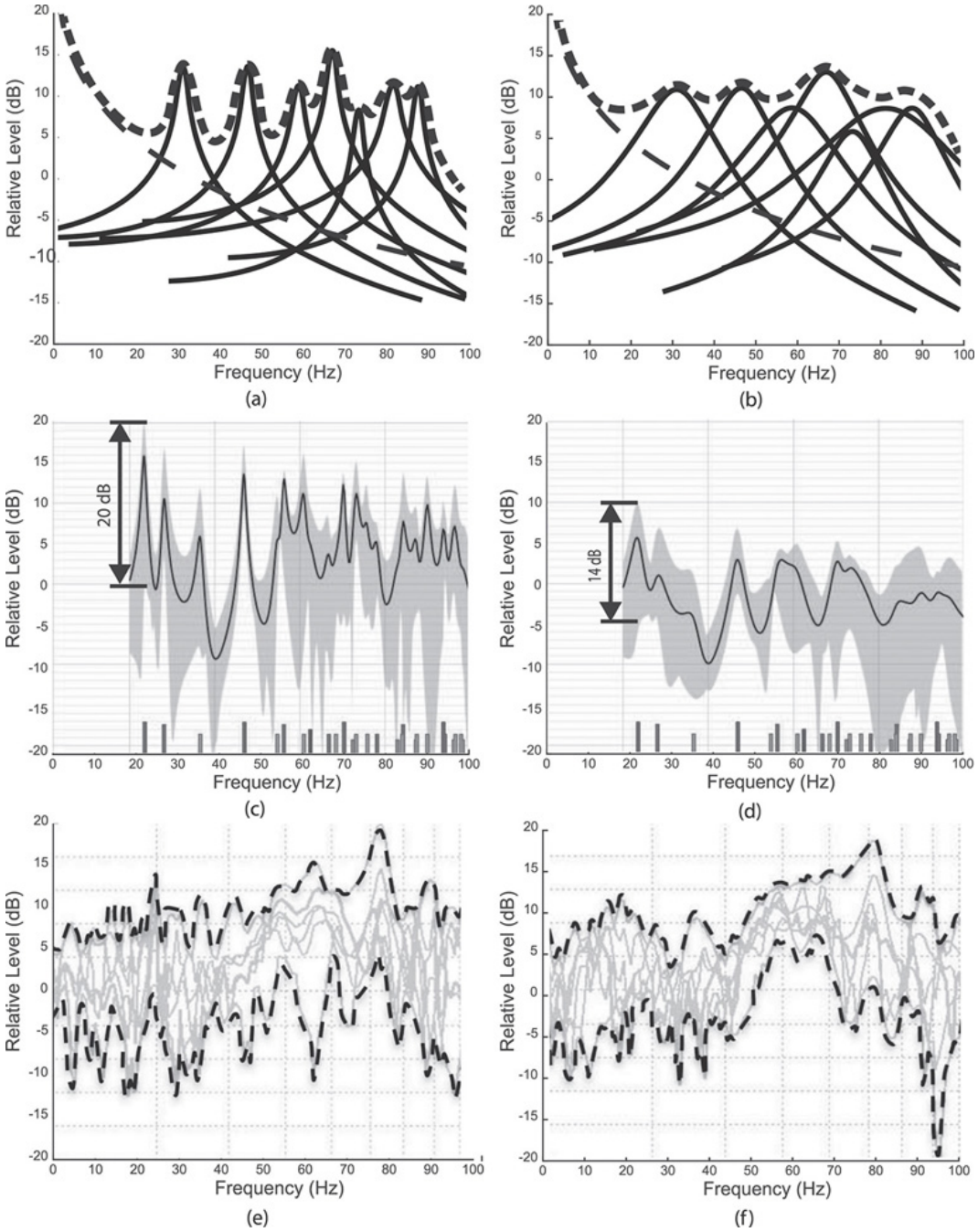


FIGURE 14.11 *The effect of adding room absorption on modes, lowering the Q of each individual mode that makes up the total response. Conceptual, modeled and measured results are shown. (a) shows generalized plot of seven modes with modest damping, (b) shows the effect of adding significant damping, with smoother overall result and more modal overlap. (c) shows modeled results (Room Eq Wizard, www.roomeqwizard.com) for rectangular room*

with sub in one corner and five seats near the center of the room, with low absorption coefficient (α = 0.1). (d) shows exact same model with high absorption coefficient (α = 0.3). (e) shows measurements in a real room for six seats with all absorption removed. In (f), the same room had soft chairs, and 4-inch absorbers covering roughly 40% of wall area. Some reduction of seat-to-seat variance in a real room can be seen in (f). The effect is limited at lowest frequencies by lack of effective absorption and at highest frequencies by high modal density (increasing the overlap by adding damping thus has little effect). ⏎

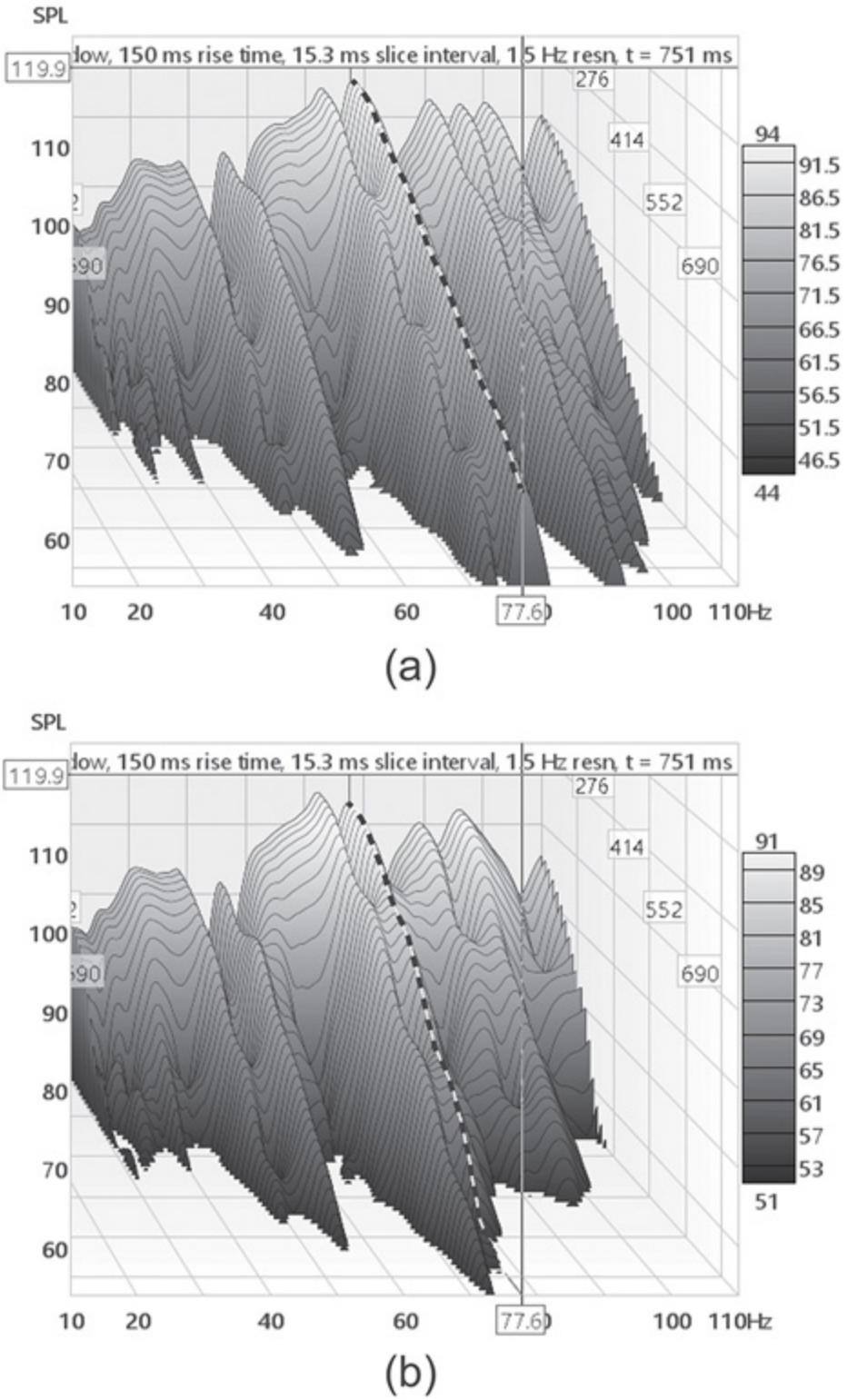


FIGURE 14.12 *Measured waterfall plots of same room at same location with minimal absorption and with "typical" absorption. (a) shows no absorption condition, whereas (b) shows with*

absorption. Decay over time of one mode at 76 Hz is highlighted, and shorter decay with absorption can be seen. Window length 500 ms, time range 750 ms. ⏎

Several important observations can be made about these diagrams:

- In Figure 14.13(b) the standing-wave ratio of the first-order mode is only 9 dB. This indicates that the walls are absorbing substantial energy at this frequency: calculated to be 23.6 Hz for the 24 ft (7.32 m) long room. Energy lost in the boundaries causes the standing wave (constructive interference) peaks to be lower and the (destructive interference) minima to be higher. Further confirmation is in the shape of the sound pressure distribution, smoothly undulating, without a sharp null as shown in (a). The spatial Q of this resonance has been reduced compared to the second-order mode, and seat-to-seat variations in bass are reduced.
- The second-order mode exhibits a larger max/min spread of 14 dB, indicating that the walls are more reflective at this frequency. The curves also exhibit sharper, albeit not sharp, dips. This mode has a higher Q than the first-order mode.
- The minimum for the first-order mode is not where it should be according to simple room dimensions—at the halfway point down the length of the room. It is shifted by about 2 feet (0.6 m) towards the wall having a door. The phase-shifted (delayed) reflection from that reactive wall makes the room "acoustically" longer than the physical dimension, and that the extension is at the end where the door is located. A boundary that moves is a membrane absorber, absorbing a portion of the energy falling on it, and reflecting the remainder with a phase shift. In this case strong vibrations were felt in the wall at 21 Hz. This is true for all membrane, or diaphragmatic, absorbers and it was a feature that was deliberately used by one of the authors in designing membrane absorbers for the NRCC listening room many years ago,

simultaneously damping modes and moving bothersome nulls away from the prime listening locations, (Toole, 1982, appendix). In this case, at this particular frequency, the phase shift has the same effect as moving the wall by some distance beyond the physical location. If this is so, there should be a corresponding lowering of the frequency at which the resonance is observed. Measurements confirmed that the true resonance frequency was 21 Hz, substantially lower than the predicted frequency of 23.6 Hz.

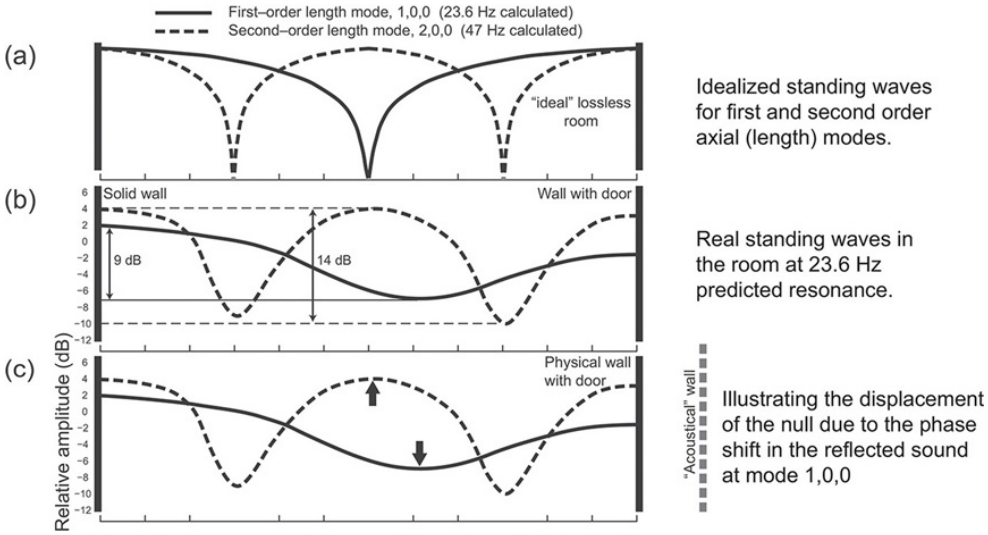


FIGURE 14.13 *(a) The stylized standing waves that might occur in a lossless, perfectly reflective, room. (b) and (c) show measurements of sound level as a function of distance along the length of a real room, measured at the frequencies of the first- and second-order modes (1,0,0 and 2,0,0). (b) shows the difference between the maximum sound level and the minimum sound level (the standing-wave ratio) for each of the modes, indicating that there is more acoustical absorption (damping) at the lower frequency. (c) shows the predicted locations for maxima and minima for the modes (the fine vertical lines at quarter- and half-wavelength distances from the boundaries). The measured pressure maxima and minima are displaced from their anticipated locations. The discrepancy is large for the first-order mode*

and small for the second-order mode. See text for explanation. Adapted from figure 8.8 in Toole (2018). ⏎

- The second-order mode (2,0,0) nulls and maximum are all very slightly displaced towards the wall with the door. However, the movement is small, indicating that at this higher frequency (47 Hz) the wall with the door is substantially less absorbing than it is at the lower first-order mode frequency.

  The simple explanation for all of this is that in spite of the solid construction of the room, the stiffness of one end wall of this room was reduced by the presence of a door in the center of it. The wall became more flexible, with possibly more mass, at very low frequencies. The performance at higher frequencies—above about 50 Hz—was unchanged. Does this matter? Not really. Low-frequency absorption is a good thing, damping the mode. The small frequency shift is normally of no consequence. It is of interest to note that, in the spatial domain, as plotted in dB, the peaks are broad and the dips are narrow. Attenuating a peak is beneficial to more listeners than elevating a dip, although damping achieves both.

The energy lost to the room-boundary membrane-absorption process can be considerable, resulting in a 2 to 3 dB reduction in amplitude over a significant frequency range, as can be seen in Figure 14.14(a). Obviously the thick, heavy, wall constructions necessary for good sound isolation aggravate low-frequency resonances, as do room walls and/or floor constructed of heavy masonry or concrete. Figure 14.14(b) shows considerably more absorption from optimally designed and located low-frequency absorbers can be. The overall trends in absorption can be seen.

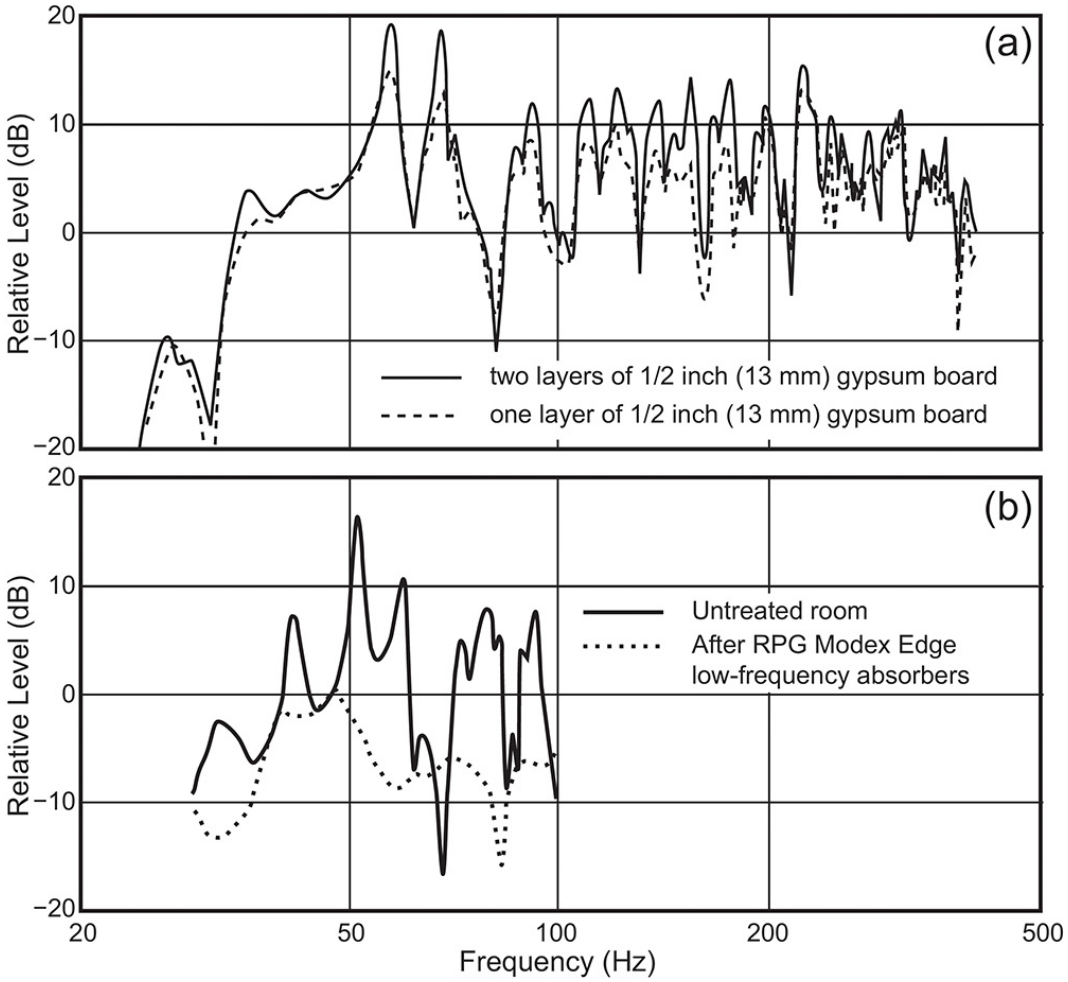


FIGURE 14.14 *(a) Measurements made with a subwoofer placed in a front corner, and a microphone located at the prime listening location in a 3000 ft^3 (85 m^3) listening room. The solid curve shows results with two layers of gypsum board and the dashed curve shows results for a single layer of gypsum board mounted normally on 2 × 6 inch (50 × 150 mm) wooden studs. (b) An example of effective damping provided by some commercial low-frequency absorbers. Courtesy RPGinc.com.* ⏎

With energy disappearing into "traps" the loudspeakers will work harder to generate sound levels, but the perceived sound will be less resonant—there is a trade-off involved. Narrow band, tuned, membrane absorbers are a logical solution when only a few modes are problematic, placed so that they

address the offending modes (use standing-wave patterns like those in Figure 14.4 to find the high-pressure regions). There are several suppliers of low-frequency absorbers, most of which are broadband, and some of them are not effective at very low frequencies, so look for realistic specifications. Broadband absorption is useful in some situations, just as narrow band is in others.

Alternatively, active bass absorbers are available. These resemble powered subwoofers but they have microphones and control electronics that turn them into absorbers of bass, not producers of bass. With electronics involved they can be very flexible in what they absorb. Passive or active, all absorbers work, especially if optimally located. Refer to OR Chapter 14 for more information on active absorbers.

> Are resonant modes eliminated by adding absorption? No, but they are acoustically damped to reduce the height of the peaks and the depth of the dips, reducing "booms" and seat-to-seat differences. Broadband bass traps remove energy at frequencies other than those involved with the resonant modes; the room will be less reflective overall, which may or may not be desirable. Narrow-band, tuned membrane absorbers can address the problem modes in relative isolation, as can some of the following solutions. In general, though, low-frequency absorption is desirable, as it will benefit any additional remedies that may be selected.

14.5.2 Deliver Energy to the Modes and Reduce the Coupling of That Energy to the Listener by Optimizing the Listening Location

Here we are modifying one of the modal coupling factors, Ψ_R in the Equation 14.2 numerator. Moving closer to or further from a null in the standing wave of a bothersome mode changes the sound pressure at the

ears. This can be a solution for a solitary listener, or in some situations, a small number of listeners in the right locations. See Figure 14.15 for a real example with frequency and time-domain data. This was in Floyd's "classical" music listening room some years ago. The room had a deliberately irregular shape to inhibit the development of strong room modes, and whether it did or not it would cause the damping to be better distributed among the modes as discussed in Section 14.4.2. In this case the energy in resonances was concentrated in a second-order length mode between the only two parallel walls of significant area. This was a room with one problem resonance, but it was a monster. Bass transients boomed outrageously and certain organ pedal notes were overwhelming.

Numerous large view windows and an aversion to unattractive objects led to the desire for alternatives to low-frequency absorbers. This situation is described in Figure 14.15 where "positional equalization" is shown, i.e. finding a seating location where the ears were not strongly coupled to the offending standing wave.

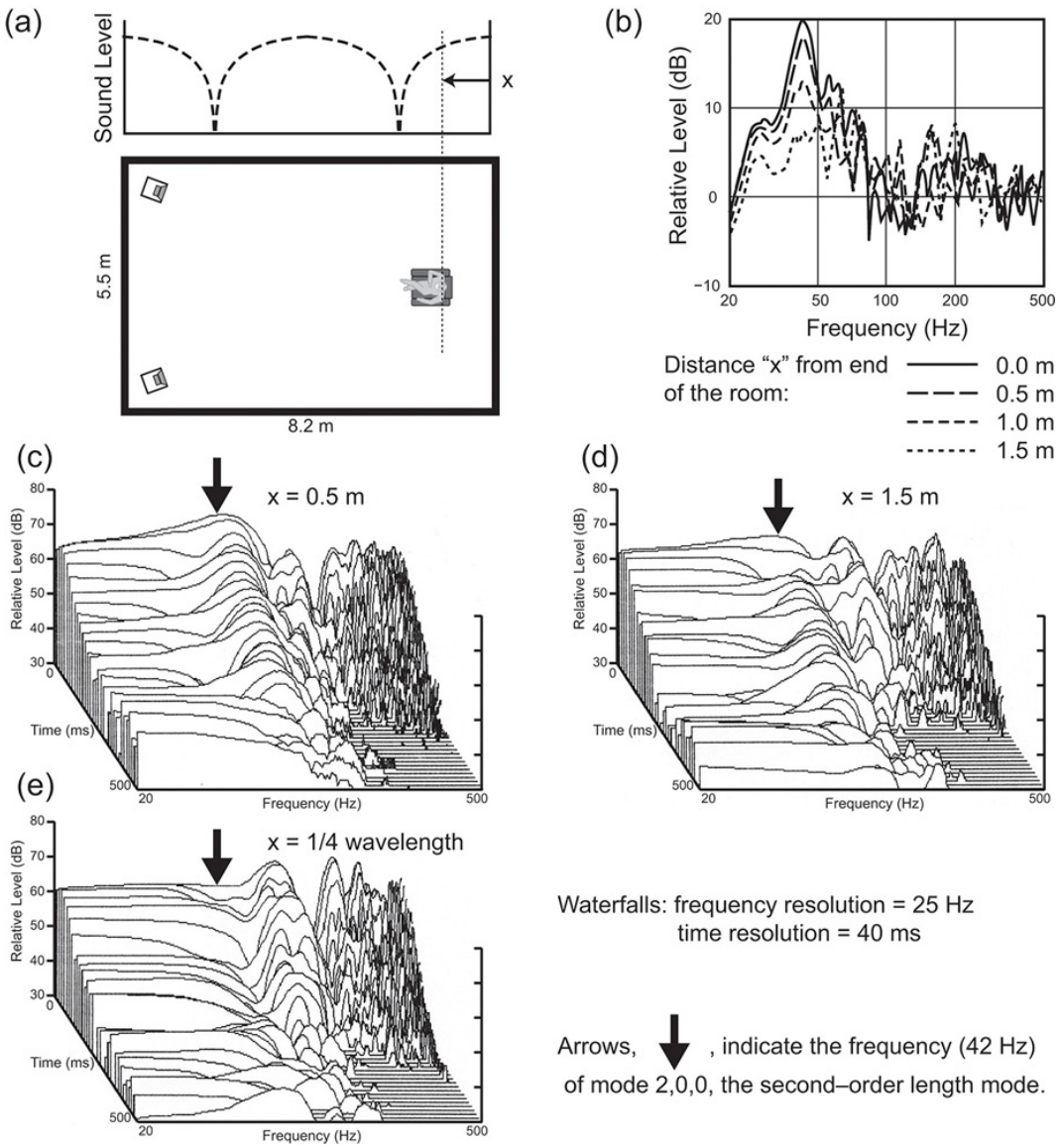


FIGURE 14.15 *Measurements made in a large (7770 ft^3/220 m^3) living/listening room. (a) The floor plan showing the second-order length mode (2,0,0) with 1/4-wavelength minima at 2 m from the end walls. (b) High-resolution frequency responses at different distances from the right end wall showing the reduction of the resonant peak as the minimum is approached. (c), (d) and (e) show waterfall diagrams at two of the locations in (b) and at the 1/4-wavelength minimum. The waterfalls have frequency resolution of 25 Hz, which corresponds to a time resolution of 40 ms. Adapted from Toole (1990; 2018, figure 8.10).* ⏎

In Figure 14.15(c) it can be seen that the reduced frequency resolution of the waterfall plot has smoothed the 42 Hz resonant peak seen clearly in (b). However, the reward is that we are able to see the prolonged ringing as time moves forward in the waterfall. A sharp resonant peak in the steady-state room curve correlates with significant ringing. The bass was intolerably boomy within 0.5 m of the wall. In (d) the listening/measurement point has been moved towards the null and, as seen in (b) the resonant peak seen in the back curve is attenuated. This is accompanied by a time-domain response that drops very quickly by about 12 dB and then decays. The bass boom was greatly reduced, and the bass sounded substantial and tight.

Going to an extreme (e) shows behavior in the 1/4-wavelength null. There is no evidence of the resonance at all in either the frequency or time domains but, subjectively, there was insufficient bass. So, the ideal solution is to attenuate the resonance by an optimum amount, but not to eliminate it. The best sounding condition was x = 1.5 m, a slight bass rise, as shown in (b).

This was a totally passive solution—no modifications except that the chair ended up in an inconvenient location 5 feet out from the back wall, leading to the desire for a better solution. It should be noted that however interesting waterfalls are, they add no information that cannot be inferred from the frequency response curves. As the curve flattens, the ringing reduces. Room resonances behave as minimum-phase systems. See Figure 9.11 for additional perspective on waterfall measurements.

When there is more than one seat in the room, it is a bit more complicated but similar principles can be applied. Figure 14.16 shows such a scenario. Placing listeners at the null lines of first- or second-order room modes (1,0,0, 2,0,0, 0,1,0, 0,2,0) is to be avoided if possible, since it will tend to result in larger seat-to-seat variations. It may be unavoidable, however, for example having a seat along the centerline of a home theater. In the example, the subwoofer is located such that the first-order width axial mode running down the centerline of the room was not excited (showing only a dashed line where it would be).

Did moving the seats affect the modes? No. In the first example an annoying bass boom was attenuated, but only for one listener. In the second example the excessive nulls were avoided at the ears of the listeners. There was no change to the overall reflective characteristics of the room, which, in this case, was not the problem. This method of optimizing modal response is probably the simplest, though moving seats around in a dedicated listening room is not trivial.

14.5.3 Reduce the Energy Delivered to a Bothersome Mode by Optimizing the Loudspeaker/Subwoofer Location

Here we are modifying one of the modal coupling factors, Ψ_S in the Equation 14.2 numerator. Moving the sound source away from a room boundary closer to a null reduces the energy transfer to that mode (Figure 14.6). This is almost the same as is discussed in the previous section, with the difference that the attenuation of the room mode is experienced everywhere in the room. The optimization can work for a single listener, or a few, but others take their chances. If neither moving seats nor moving subwoofers solves the problem, we reach for more powerful methods.

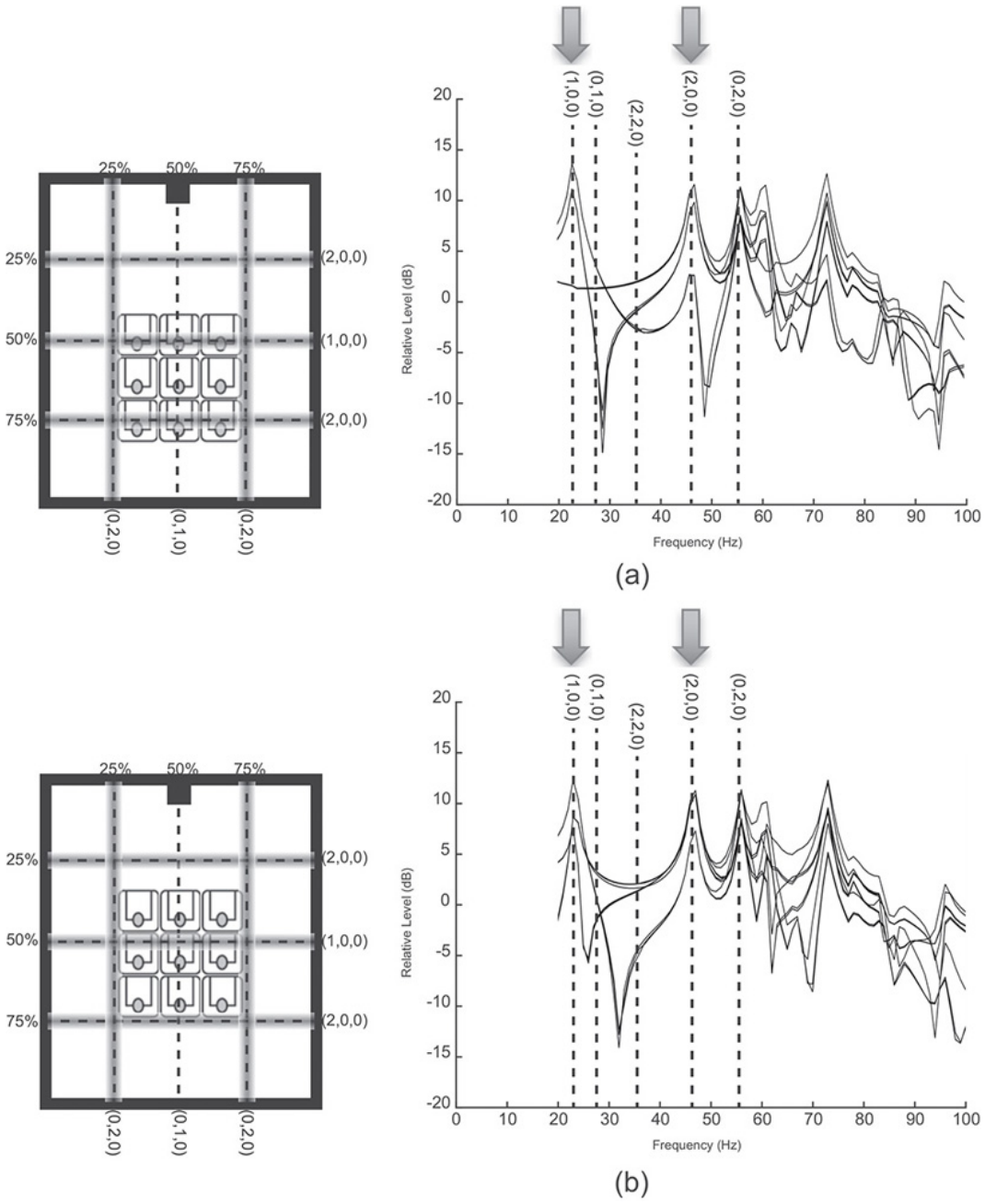


FIGURE 14.16 *Effect of moving seating area slightly to avoid axial nulls. Shaded lines show mode nulls that are active for the front center subwoofer shown. Example room is 24 × 20 × 9 feet (730 × 610 × 270 mm), with subwoofer in front wall midpoint. (a) shows reference seating layout, where front and rear rows have the listener's heads located on the first and second axial length mode null lines (1,0,0 and 2,0,0). In (b) the seats have been*

moved forward 1.5 feet (50 cm) to avoid the nulls. A modest reduction in seat-to-seat variance can be seen at those frequencies. The arrows show the first- and second-length modes. ⏎

Would the mode be eliminated? No, but the energy in the problematic mode could be reduced to a useful degree for the listener in the prime location and possibly for other locations. But there would be substantial seat-to-seat variations. The room acoustics (Equation 14.2 denominator) are again unchanged.

14.5.4 Reduce the Energy Delivered to Bothersome Modes by Using Combined Coupling of Two Sources to Mode

Section 14.3.1 described how modal coupling works. Now we will see how that can be useful in the real world. Figure 14.17 shows the basic methods of manipulating the supply of bass energy to low-frequency modes.

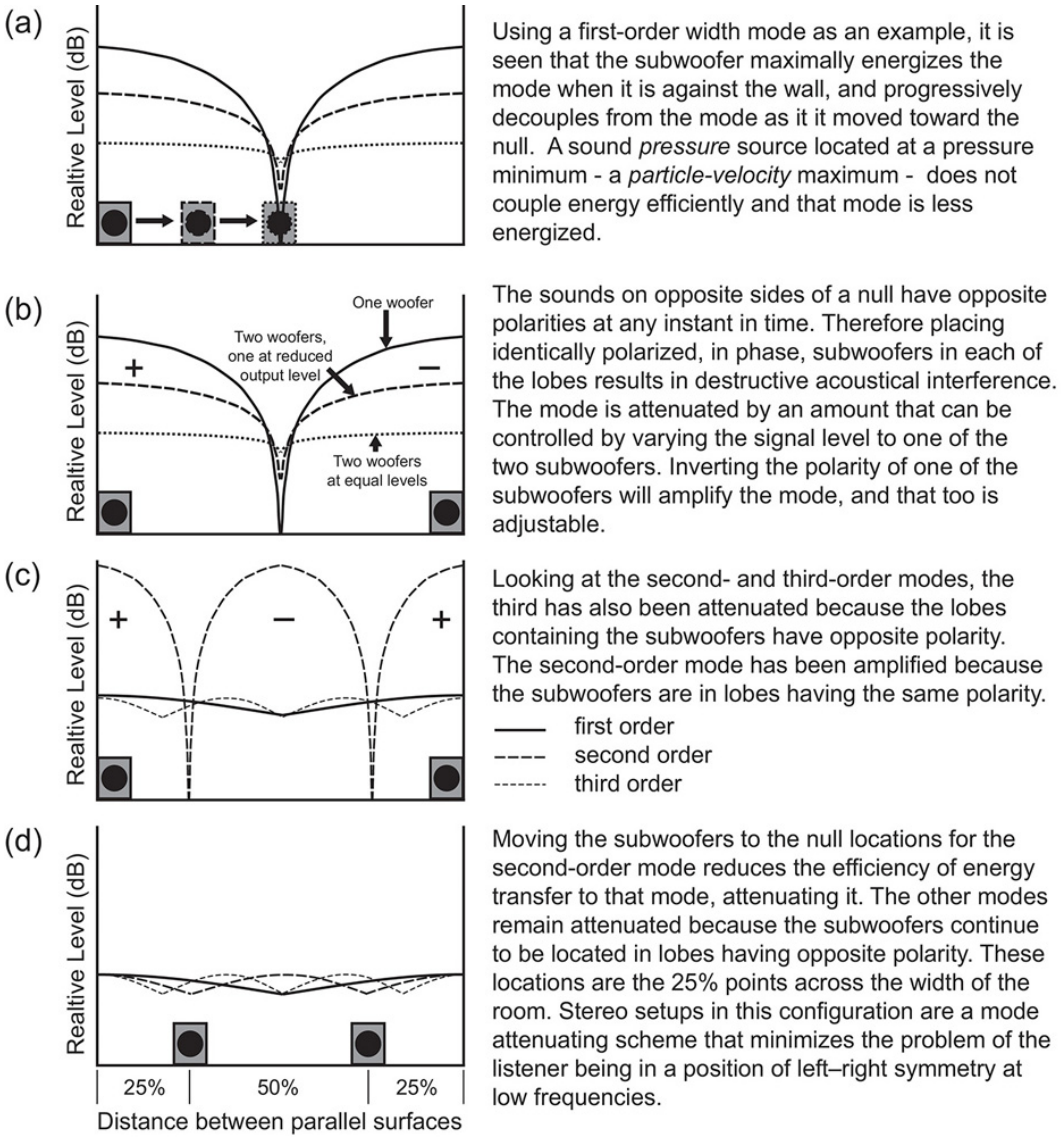


FIGURE 14.17 *A simplified, exaggerated illustration of how mode manipulation using multiple sources works. (a) Subwoofer location is used as a means of controlling the amount of energy supplied to a bothersome mode. (b) Taking advantage of polarity reversals in the adjacent lobes of standing waves, two subwoofers are used to vary the amount of attenuation of a specific mode, potentially achieving cancellation. (c) presents a more realistic view by including three orders of width modes. The odd-order (first and third) width modes are attenuated because the subwoofers are in the opposite-polarity lobes of these modes, but the second-order mode is amplified because the subwoofers are in*

lobes exhibiting the same polarity. (d) shows a combination solution, in which approach (a) addresses the second-order mode and approach (b) addresses the first- and third-order modes. ⏎

In the case that two subwoofers are placed with opposite polarity so as to cancel each other for a given mode, Equation 14.2 numerator (coupling factors) might be stated as:

$$\Psi_S\Psi_R = (\Psi_{S1} + \Psi_{S2})(\Psi_R) = ((-1) + (+1))(\Psi_R) = 0$$

Perceptive readers will have noticed that conventional arrangements of stereo loudspeakers will result in the cancellation of the first-order width mode because most bass in stereo recordings is essentially monophonic (therefore near perfect cancellation). The attenuation works, thereby liberating the symmetrical axis of the room for the stereo seat, which is no longer in a profound null. Taking it further, those listeners who can place their loudspeakers at the 25% positions will find three width modes attenuated. It is doubtful that many people ever thought about basic stereo as a room mode–manipulating scheme. It has been working for us for decades and does so very well because most bass in stereo recordings, and all bass in LPs, is monophonic. Note that in Figure 14.17 we are assuming a generalized background of modal activity (Helmholtz resonance and adjacent modes) that prevents the response from going to 0, i.e. $-\infty$ dB when modes are cancelled.

Another of Floyd's listening rooms also suffered from an enormous bass boom. It was quickly traced to a combination of a first-order width mode and a second-order length mode having almost identical frequencies, both being well energized by a single subwoofer in a corner. Because of the large view windows, large bass traps were not an option. It was taken as a challenge to solve the problem without degrading the appearance of the room. The floor plan is shown in Figure 14.18(a). Note that this is not a rectangular room, but perhaps it behaves similarly enough to apply mode cancelling techniques. Employing the logic of Figure 14.17(b) the single

sub in the corner (position 1) was replaced with two identical subs located in positions 2 and 3. The subs, driven with identical signals, were in opposite polarity lobes of the width mode and also of the length mode, so for both modes the *combined* modal coupling factor was very low and mode energy substantially attenuated. The resonant peak was attenuated, and one of the modes was truly cancelled as is demonstrated by the narrow destructive-interference dip in Figure 14.18(b). The other improvements were fortuitous, and never analyzed, following the "leave well enough alone" principle. The good behavior can also be seen in the waterfall diagrams, as can evidence of energy sharing between the two closely spaced, overlapping, resonances. In Figure 14.18(d) it is clear that the modes are merged, and with both subs operating the higher frequency one is seen to decay very quickly, and the lower frequency one is greatly attenuated before decaying more slowly. The latter could have resulted in some pitch shifting in the very early part of the decay, but no listeners appeared to hear it; the bass sounded satisfyingly smooth and tight. There was no equalization. Again, the trustworthy evidence of good behavior is in the steady-state room curve; the waterfall diagrams provide decorative confirmation.

Floyd presented this data at an AES conference (Toole, 1990). It may have been the first application of deliberate mode cancelling in an audio room-acoustics application.

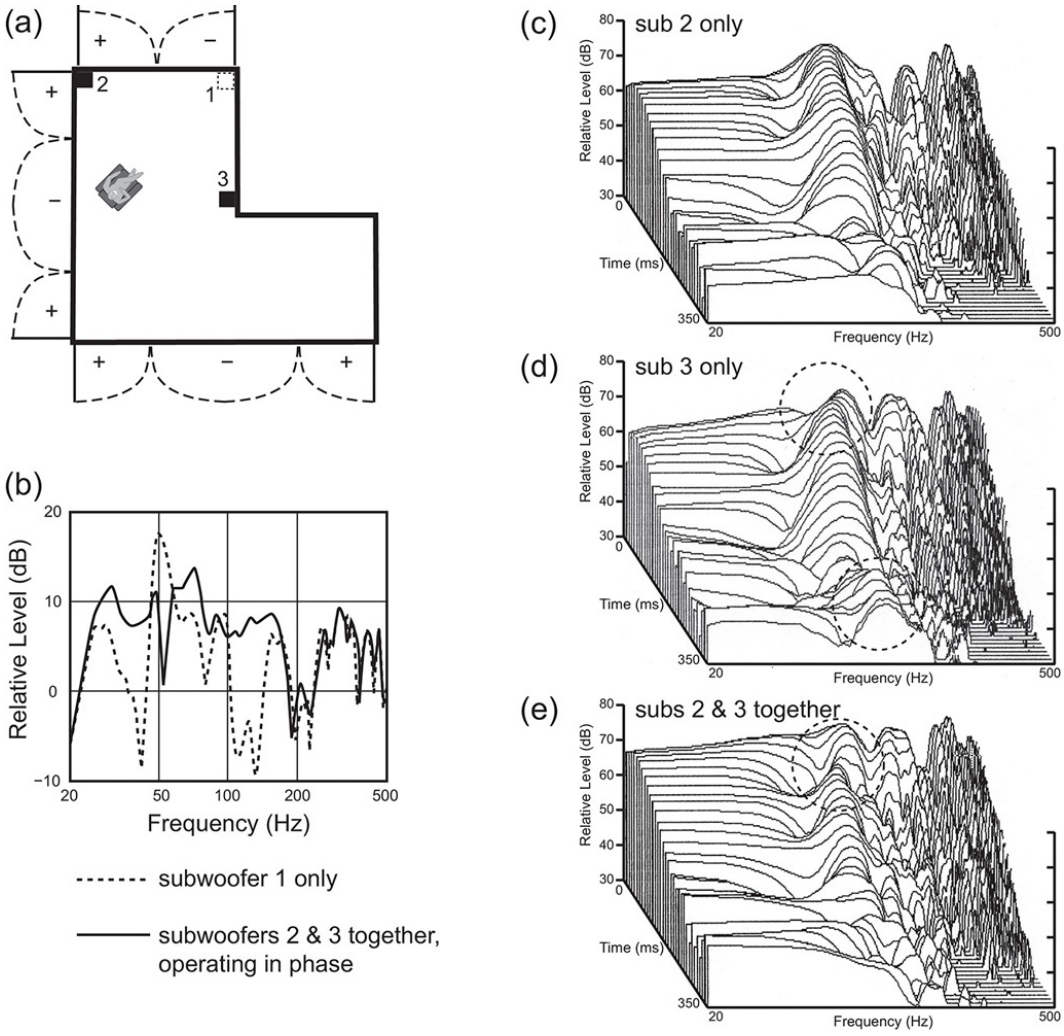


FIGURE 14.18 *A family room attached to a dining area and kitchen in an L-shaped space. (a) shows that there were three axial modes at similar frequencies, all of which were strongly energized by a single subwoofer located in the corner behind the video display—subwoofer 1. (b) shows (dotted curve) that there was a powerful resonance around 50 Hz, and large holes in the bass response above and below that frequency. Inspection of the polarities in the standing wave lobes suggested that a pair of subwoofers, one in a lobe of each polarity, should alleviate the problem. This was conveniently possible, at locations 2 and 3, and the solid curve in (b) was the result, with no equalization. The prominent resonant peak has been transformed into a narrow acoustical interference dip. (c) shows the waterfall at the prime*

listening position for subwoofer 2 only, (d) shows one for subwoofer 3 only, and (e) shows one for both subwoofers. In (c), the energetic ringing is obvious, and it appears that, with time, the ringing skews downward in frequency. In (d) it becomes clear that the reason is that there are at least two closely adjacent modes that are revealed because they have different decay rates (amounts of damping). See what appears to be a quite discrete downward frequency shift in the top circle, from one rapidly decaying mode to a more energetic lower-frequency one, leaving a single resonant tail in the bottom circle. In (e) it is clear that, with both subwoofers operating, the aggressive ringing is substantially attenuated. This is seen as a rapid 12 dB drop in amplitude in the very early decay before low-level ringing disappears into room noise. Evidence of the two modes can still be seen early in the decay. (a) and (b) are adapted from Toole (1990, figures 14 and 15). This is figure 8.14 in Toole (2018). ⏎

Were the targeted modes eliminated? As for the room, the modes are unaffected, but for the sound playback system, they were neutralized. The benefits were clearly audible in all of the seats in the room.

14.5.5 Use Standardized Subwoofer Configurations for Generalized Multiple Subwoofer Mode Manipulation—Basic

Not long after the authors started working together, we thought that mode manipulation should be able to work in a more generalized sense, not just for one mode in a specific custom room installation. Todd first tackled the case of simple rectangular rooms, and after substantial analysis, involving simulations using from one to 5000 subwoofers (acoustic wallpaper?) concluded that two or four subwoofers—identical, and identically driven—created a central zone in which seat-to-seat variations were usefully reduced

(Welti, 2002a, 2002b). This means that equalization can benefit several listeners.

The optimized subwoofer configurations discussed in this section have been shown to work well *in general*. There are of course no guarantees; they are "rule of thumb" suggestions. If room dimensions, subwoofer positions and seating positions are known, more powerful methods are available in Section 14.5.6.

Before going any further, let us introduce more formal terminology for some objective metrics. The most important by far is Mean Spatial Variance (MSV). Other similar metrics have been used by other researchers. Lower values for MSV indicate less variation of frequency response from seat to seat, which is the goal. Low-frequency efficiency is expressed by Mean Output Level (MOL), quantifying a mean output level from 20 to 40 Hz, including correction factor for the number of subwoofers ("bass-per-sub", if you will). Finally, we have variation of spatial average (VSA), which describes how flat the response at each seat is on average. Refer to Table 14.2.

Figure 14.19 illustrates the underlying logic of the process in a typically dimensioned room. Figure 14.19(a) shows that a single subwoofer in a corner energizes all of the horizontal-plane modes in the room producing null (anti-node) lines as indicated in the floor plan. The MSV was calculated for a square area in the center of the room, indicating the potential seat-to-seat variations in bass level. With all the modes active, the MSV was high. The effect of moving the single subwoofer from the corner to a mid-wall position is shown in Figure 14.19(a)—the first-order width mode (0,1,0) at 28 Hz was attenuated because the subwoofer was located at a pressure minimum. Because the one of the nulls running through the seating area was removed (the dotted line) the MSV was reduced, but only slightly.

In Figure 14.19(b) the addition of a second identical in-phase subwoofer at the center of the opposite wall cancels the first-order length mode (1,0,0) at 23.5 Hz (as shown in Figure 14.17(b)) and the accompanying odd-order mode (3,0,0) near 70 Hz. Eliminating these nulls in the listening area results

in a dramatic reduction in the MSV. However, the two woofers now amplify the second-order length mode (2,0,0) at 47 Hz giving it *more* amplitude. The gain in the second-order width mode (0,2,0) at 56 Hz appears to be the result of having subs at both ends of the room, symmetrically driving this mode from the central high-pressure region. But, because the pressure nulls of second-order modes are at the 25% and 75% locations between room walls, they miss the designated listening area. The amplitude gains in modes 2,0,0 and 0,2,0 help boost the bass efficiency. However, the MSV is attractively low, and equalization to reduce the resonance peaks in fact gives the system increased headroom at those frequencies, so this is a viable option. Here we are limiting ourselves to subwoofer locations along the walls. See OR Chapter 14 "Toole configuration" to see an example of what can be done without this limitation.

TABLE 14.2 Three metrics used to qualify frequency response of a given subwoofer configuration across multiple seats. All values calculated in dB (resulting in units of dB^2 for variance metrics). Typically, the maximum output capability of subwoofers decreases at lower frequencies, which determines the ultimate SPL that this system can achieve. For this reason, we calculate the MOL for the 20 to 40 Hz range. These metrics are for the purpose of relative comparisons of subwoofer configurations only. Refer to OR Chapter 14 for discussion about spatial averaging and why we do it in dB. ⏎

Metric	Equation	How to calculate	Meaning
Mean Spatial Variance	$MSV = mean_f\,(var_S\,(P_{dB}\,(f, S))(dB^2)$	*At each frequency*, take the variance of frequency response P across all seats S. Then take mean of these variance values to get a single number.	Variation of subwoofer response from seat to seat. Low values are good.

Metric	Equation	How to calculate	Meaning
Mean Output Level	MOL= $Mean_r(mean_S(p_{dB}(f_{20\text{-}40}, S))-10log(N)(dB)$	Take the mean frequency response P for all seats S and all frequencies from 20 to 40 Hz. Subtract 10log(N) to account for number of subwoofers N, to get efficiency metric.	Low bass efficiency. High values are good.
Variance Of Spatial Average	$VSA = var_f(mean_S(P_{dB}(f, S))(dB^2)$	*At each frequency*, take the mean across all seats S (spatial average), then take the variance of that spatial average across frequency to get a single number.	Variation of average frequency response across all seats. i.e. flatness. Low values are good.

In Figure 14.19(c) it is seen that adding two more identical mid-wall subwoofers puts in-phase excitation in all three pressure maxima of the second-order modes along both length and width axes. They are greatly attenuated, significantly reducing the system efficiency. But what is happening around 74 Hz? While we were attending to the lower frequency axial modes a tangential mode (OR Chapter 14) has been creeping up in amplitude and now dominates the acoustic delivery, thereby compensating somewhat for the loss of the axial modes. This is because we have placed in-phase subwoofers at in-phase regions in the standing-wave pattern for this mode and it is responding. Figure 14.20(a) shows what is happening. The seat-to-seat variations are low because this mode includes all listeners within a broad high-pressure region, and this is now, by far, the strongest mode in the room.

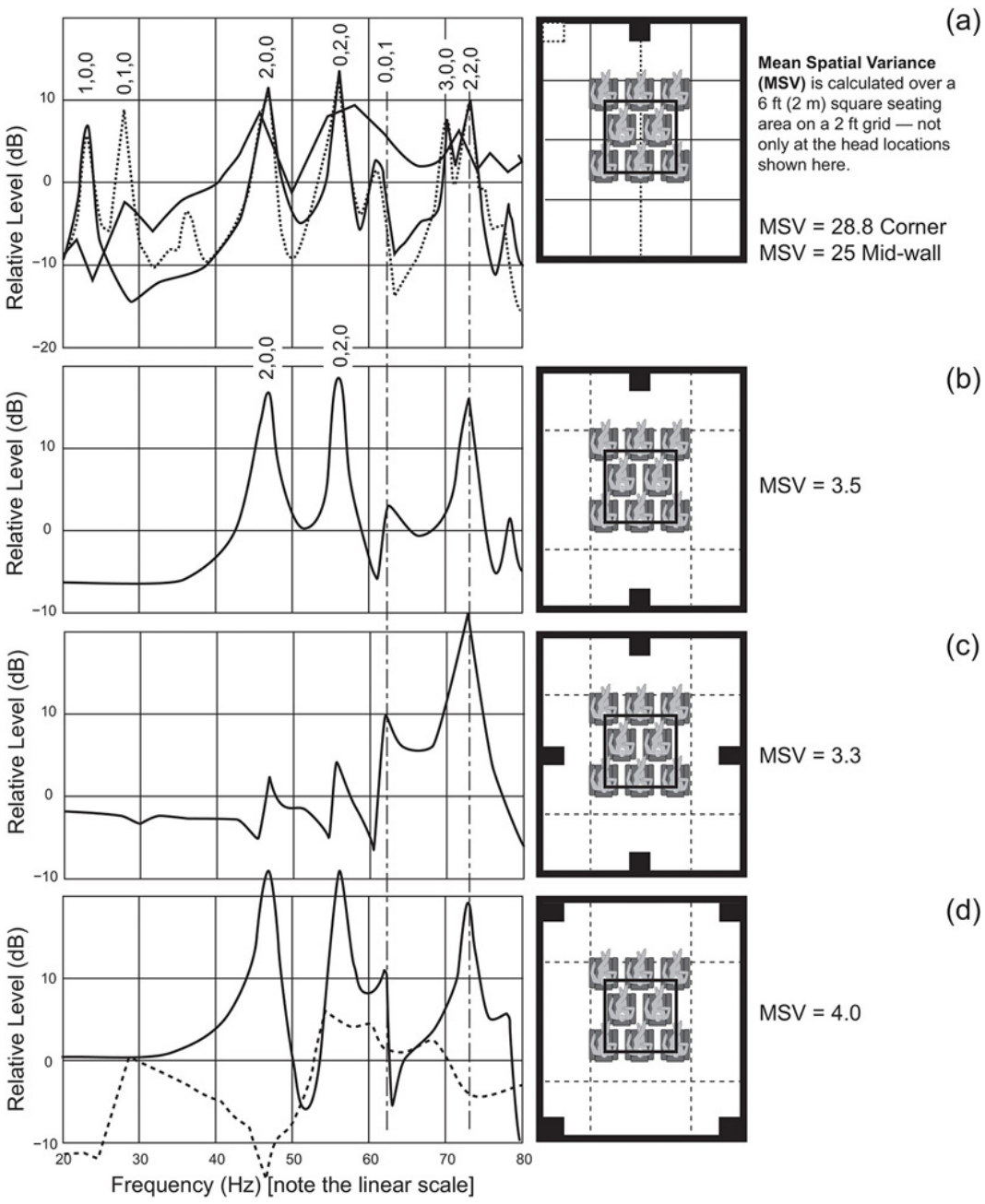


FIGURE 14.19 *Four different optimal subwoofer configurations in an example room. Plots consist of "average total" frequency response curves taken from Welti (2002a and 2002b). This is the combination of direct sound from the subwoofers and the modal energy within the room. They have been traced and vertically rescaled to conform to the standard used throughout this book. The mean spatial variance (MSV) values are from Welti and*

Devantier (2006, figure 5(c)) for the same loudspeaker configurations. ⏎

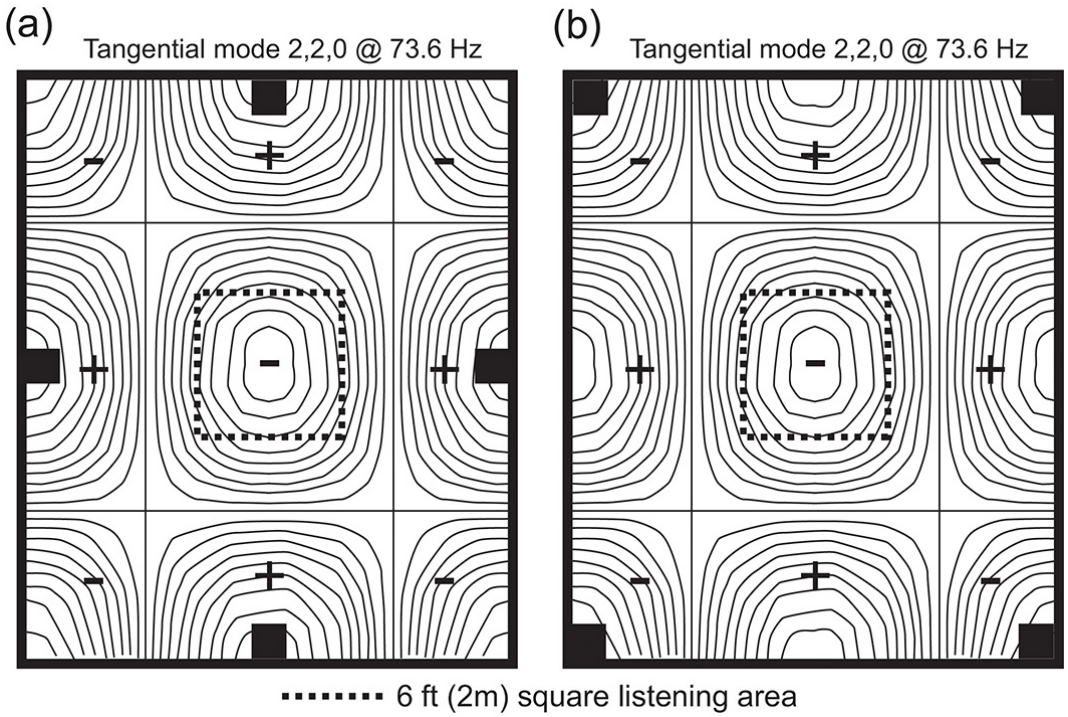


FIGURE 14.20 *(a) The pressure contours for tangential mode 2,2,0 are shown for the test room. Subwoofers in two configurations are shown to be in high pressure points that all share the same polarity, resulting in maximal stimulation of this mode. The listening area is shown to be fully within the central high-pressure region, resulting in this mode being strongly visible in the frequency response of Figure 14.19(c) and (d), and in the low seat-to-seat variation metric, MSV.* ⏎

Moving to Figure 14.19(d), it is seen that placing the subwoofers at the four corners is also a great improvement in reducing seat-to-seat variations, with a considerable increase in overall sound level compared to the four mid-wall locations in (c), but only slightly better than the two mid-wall subwoofer solution in (b). One note about wall midpoint vs. corner positions: modal nulls are fairly localized (and may not be at the actual wall

midpoint as we have seen), whereas corners are not nulls and not nearly as sensitive. As we will see in Section 14.5.6, so-called Quartal placement (FBQ and LRQ) of four subwoofers can also be a good compromise.

14.5.6 Use Standardized Subwoofer Configurations for Generalized Multiple Subwoofer Mode Manipulation—Advanced

These methods work, but there are questions. Does this analysis, based on "typically" dimensioned rooms, apply for all small rectangular rooms with different dimensions? What about the trade-offs between reducing the seat-to-seat variations and the efficiency of the resulting systems? Welti (2012) revisited this topic and expanded the investigation to include more variables. Using a custom MATLAB model, a wide range of room dimensions were analyzed and various seating arrangements, including larger more realistic listening areas positioned closer to the rear of the room. Several different subwoofer configurations were also tested. The results include a rather large compendium of plots and tables and additional data such as VSA, all of which are available in color online at www.harman.com/audio-innovations and OR Chapter 14. Figure 14.21 shows the variables in his calculations.

The results and discussion for Figures 14.22 and 14.23 are for averaged data over the entire range of rooms evaluated, at 13 × 13 feet to 30 × 30 feet (4 × 4 to 9 × 9 m), in 8-inch (0.2 m) steps, ceiling fixed at 9 feet (2.7 m). At the end of this section we discuss optimization using actual specific room dimensions, a much more powerful method.

The listening areas used here are different from that used in the preceding examples, so the MSV numbers will not be directly comparable. Given that we have already seen that moving listeners away from low-order null lines is advantageous, it is clear that some of the large seating configurations in the new study will yield higher MSVs than the smaller, more centrally located areas. Those that avoid the null lines (LCm and LRm) would be expected to show improvements over those that do not.

Figure 14.22 shows mean MSV values for these configurations. It can be seen that the small, centrally located listening areas SC and SR are advantageous, along with the Large Rear modified (LRm) arrangement that avoids the second-order null lines. This shows up very clearly in the averaged values along the bottom of the table. If one were to choose the best of the listener arrangements, SC, SR and LRm, the averages along the right side of the table indicate an advantage for the 4M loudspeaker arrangement, but the spread among them is small compared to other numbers in the table.

Figure 14.23 answers the second important question: now that one can identify listener and loudspeaker arrangements that yield superior seat-to-seat consistency, which listener and loudspeaker arrangement(s) generate the highest sound levels from a given set of subwoofers? The MOL metric gives a relative low-frequency efficiency for each configuration (averaged SPL over all seats from 20 to 40 Hz, normalized to account for the number of subwoofers used).

The answer is the same for the listener arrangements, but with respect to loudspeaker arrangements the outstanding winner for efficiency is 4C—four subwoofers, one in each corner. Not far behind are the FBQ and LRQ subwoofer configurations. The mid-wall subwoofer arrangements, which are the best in terms of MSV (Figure 14.22), do so at the expense of higher power requirements and/or larger subwoofers.

The lower-performing listener configurations are often those that embrace problematic null lines, thereby increasing seat-to-seat variations. However, the prime listening areas SC, SR and LRm are included (area-wise) in those larger configurations. So, if one of the larger areas is necessary to accommodate a large audience, one option would be to select the FBQ or LRQ subwoofer configurations, if possible. They seem to be generally good compromises between MSV and MOL. Another option would be to employ a subwoofer configuration that performs well for the smaller areas and simply accept the reality that some seats will be better than others. Perhaps the center seats are the “money” seats, and we know where they are!

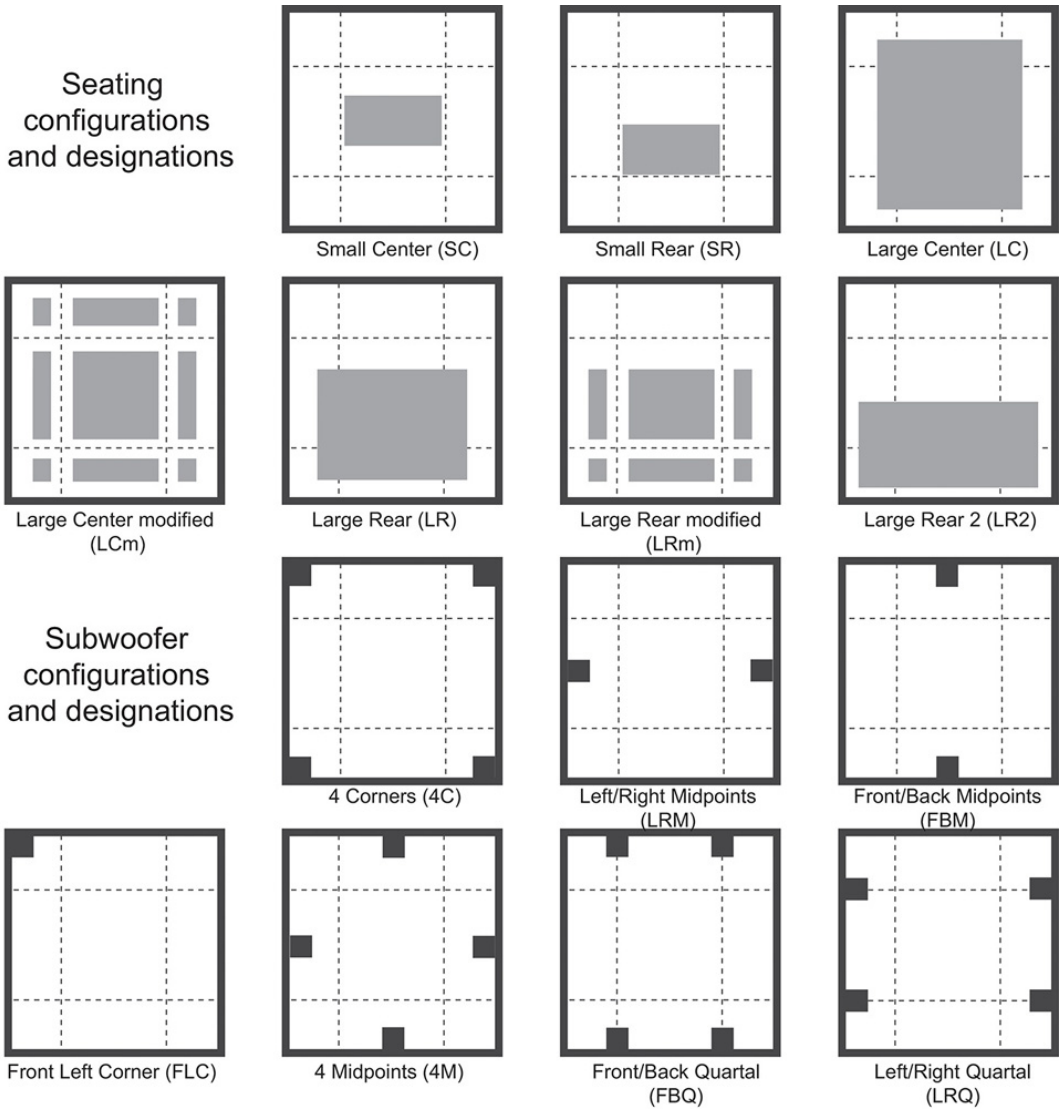


FIGURE 14.21 *The seating and subwoofer configurations used in the Welti (2012) study. As before, all subwoofers are identical, operating with the same polarity and gain. The dashed lines are the null locations for the second-order axial and tangential modes.* ⏎

A general summary of this comprehensive investigation is shown in Figure 14.24, but note that this summary averages all the analyzed room dimensions so these very general conclusions may not be valid for any one particular room.

	SC	SR	LC	LCm	LR	LRm	LR2	Ave.	
FLC	27	32	41	36	41	36	41	**36.3**	
4C	12	22	35	28	35	26	37	**27.9**	20
LRM	12	19	28	23	28	**21**	31	**23.1**	17.3
FBM	**10**	19	28	22	28	**21**	31	**22.7**	16.6
4M	11	**15**	**21**	**21**	**21**	**21**	**22**	**18.9**	15.6
FBQ	**10**	24	26	24	26	**21**	30	**23.0**	18.3
LRQ	16	18	28	22	27	22	30	**23.3**	18.6
	11.8*	19.5*	27.6	23.3	27.5	22*	30.1		

FIGURE 14.22 *Mean MSV values (dB^2) for seats within the configurations shown in Figure 14.21, listener arrangement horizontally and loudspeaker arrangements vertically. In this table, lower values indicate smaller seat-to-seat variations. The best values for each seating arrangement are in bold. The basic table is from Welti (2012, table 2). The additional numbers added outside the table boundaries by the author are, horizontally, the averages of top best listener-configuration scores SC, SR and LRm, ignoring the poorly performing reference condition: FLC. The numbers at the bottom are the average loudspeaker-configuration scores for the best listener arrangements SC, SR and LRm. Asterisks indicate the top three scores.* ⏎

	SC	SR	LC	LCm	LR	LRm	LR2	Ave.	
FLC	−6	−5	**−3**	**−3**	**−3**	**−4**	**−1**	−3.6	
4C	**−2**	**−3**	−5	−4	−5	**−4**	−7	**−4.3**	−3
LRM	−10	−9	−10	−9	−10	−9	−9	**−9.4**	−9.3
FBM	−9	−10	−10	−10	−10	−9	−10	**−9.7**	−9.3
4M	−8	−8	−9	−9	−9	−9	−10	**−8.9**	−8.3
FBQ	−4	−6	−7	−7	−7	−6	−9	**−6.6**	−5.3
LRQ	−5	−5	−7	−6	−7	−6	−8	**−6.3**	−5.3
	−6.3*	−6.8*	−8	−7.5	−8	−7.2*	−8.8		

FIGURE 14.23 *Mean MOL values (dB) for seats within the configurations shown in Figure 14.21, listener arrangements horizontally and loudspeaker arrangements vertically. In this table, higher values (the smaller negative numbers) indicate higher subwoofer efficiency over the frequency range 20–40 Hz (higher sound levels for a given signal level). The best values for each seating arrangement are in bold. The basic table is from*

Welti (2012, table 3). The additional numbers added outside the table boundaries by the author are, horizontally, the averages of the top best listener-configuration scores SC, SR and LRm, ignoring the poorly performing reference condition: FLC. The vertical numbers at the bottom are the average loudspeaker-configuration scores for the best listener arrangements SC, SR and LRm. Asterisks indicate the top three scores. ⏎

A much more powerful optimization is possible in the case that room dimensions are known or can be chosen. In that case we can look up the different subwoofer and seating configurations and try to select the best ones for that room. Further, if any two of the three parameters of room dimensions, subwoofer configuration and seating configuration are known, it is possible to optimize the third. We can also make trade-offs between MSV, MOL and VSA (though we are prioritizing MSV here). Figure 14.25 shows an example, a 16 × 23 foot (500 × 700 cm) room with a seating configuration approximating the "SR" one. We can use the front left sub configuration as a starting point. Let's assume we only have two subwoofers, and that the LRM and FBM configurations are good alternatives, and compare.

Given a design priority for bass efficiency (MOL), minimizing seat to seat variance (MSV), or flattest average response (VSA), the best option can be selected.

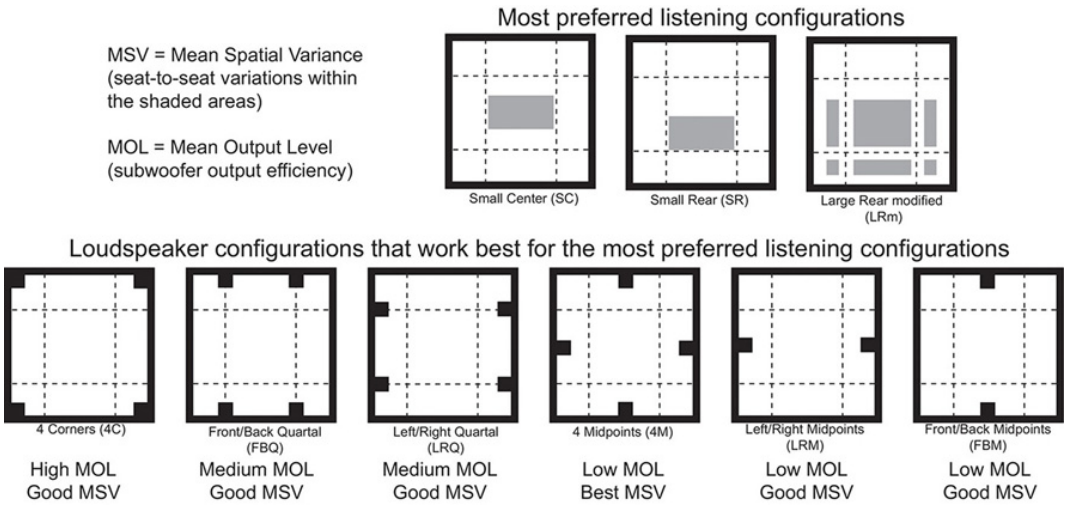


FIGURE 14.24 *A summary of Welti's (2012) investigation, assuming preferred seating configurations used. Specific room dimensions are all lumped together here, so it is very generalized.* ⏎

In general, qualifying the amount "better" for these metrics is straightforward for MOL, as it is in dB. For MSV and VSA it is a bit trickier since it is based on variance and is in units of dB^2. A suggested rough guideline, based on experience of the authors, is that a 25% decrease in the MSV or VSA values would be noticeable while a 50% or more decrease of the value would be a game changer. Also refer to Figure 14.26 for a further example of MSV vs. actual measurements. Finally, remember that seat-to-seat variation (MSV) is the only metric that cannot be improved with larger subwoofers or equalization.

14.5.7 Manipulate Modes Using Multiple Subwoofers and Computer Optimized Signal Processing—Rectangular Rooms

As has been demonstrated, rectangular rooms have advantages for passive multi-subwoofer strategies. They also are advantageous if one goes to the next level and introduces signal processing of the signals being supplied to multiple subwoofers. The idea seems to have originated with Goertz et al. (2001), who used an arrangement of subwoofers on the front wall,

generating an approximation of a plane wave that propagates to the rear of the room where a similar arrangement of subwoofers, appropriately delayed and polarity inverted, cancels the wave when it arrives. Ideally this eliminates standing waves and everyone in the room should hear good bass. It requires a perfectly rectangular room and at least four subwoofers to work well. The scheme was called double bass array (DBA).

Santillán (2001) demonstrated that a system of 16 subwoofers arranged in rows in the walls could establish a large useful listening area in a rectangular room. The installation would only be practical in a dedicated installation, and the implementation is not simple.

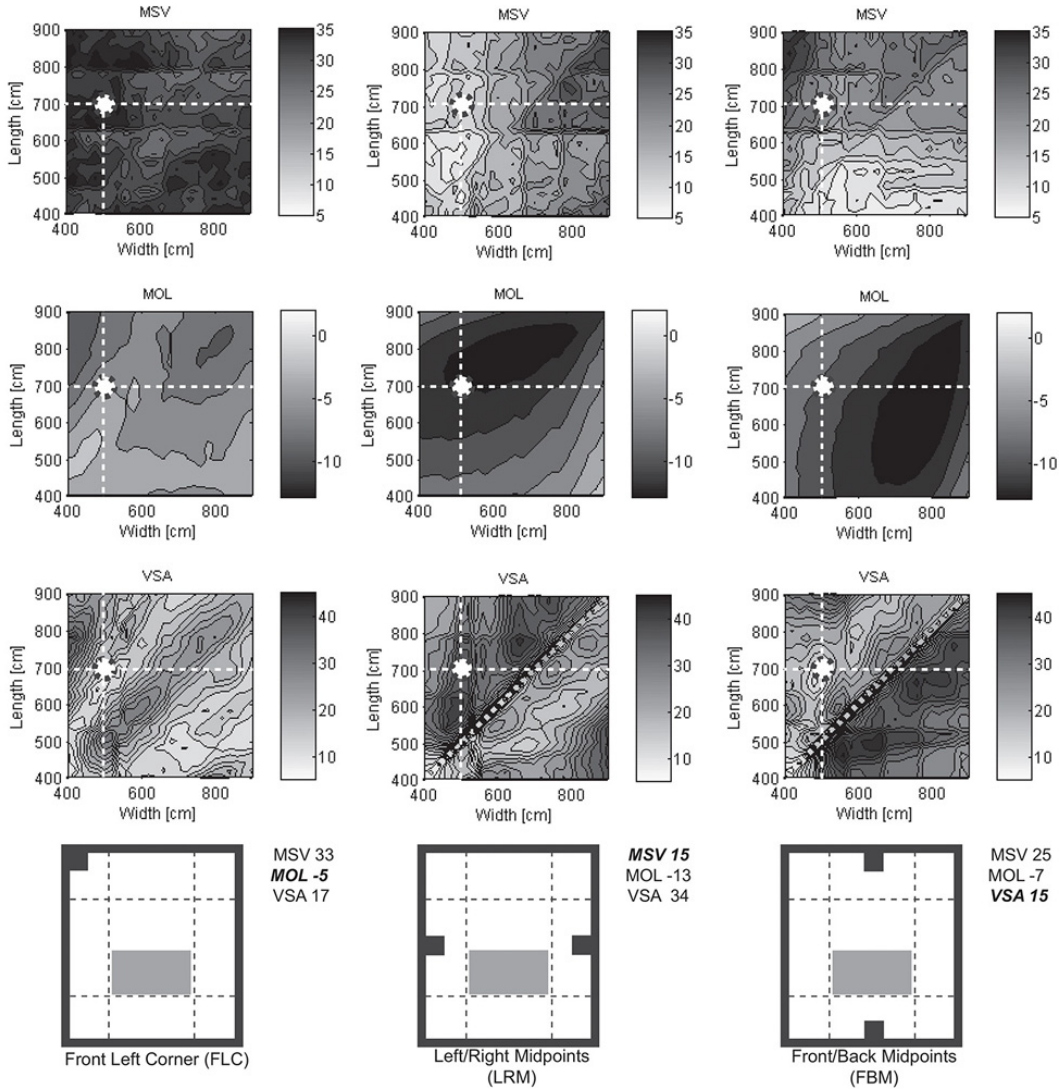


FIGURE 14.25 *An example of how to use the Welti (2012) compendium to choose between two optimized subwoofer setups for a 16 × 23 foot (5 × 7 m) room with seating similar to the "SR" configuration. Contour plots show each metric for a range of room dimensions (the horizontal and vertical axes). Note that seating areas scale with room dimensions in these simulations. In all cases, lighter areas on the plot corresponds to preferred metrics values, i.e. lower for MSV and VSA, and higher values for MOL. Metrics are summarized at bottom. Bold values are the better ones for the three alternatives.* ⏎

Celestinos and Neilsen (2008) developed the controlled acoustic bass system (CABS), a DBA scheme that could be retrofitted into an existing rectangular room. They examined the process in models and real rooms and demonstrated that it works well, yielding reduced variations in bass over much of a room.

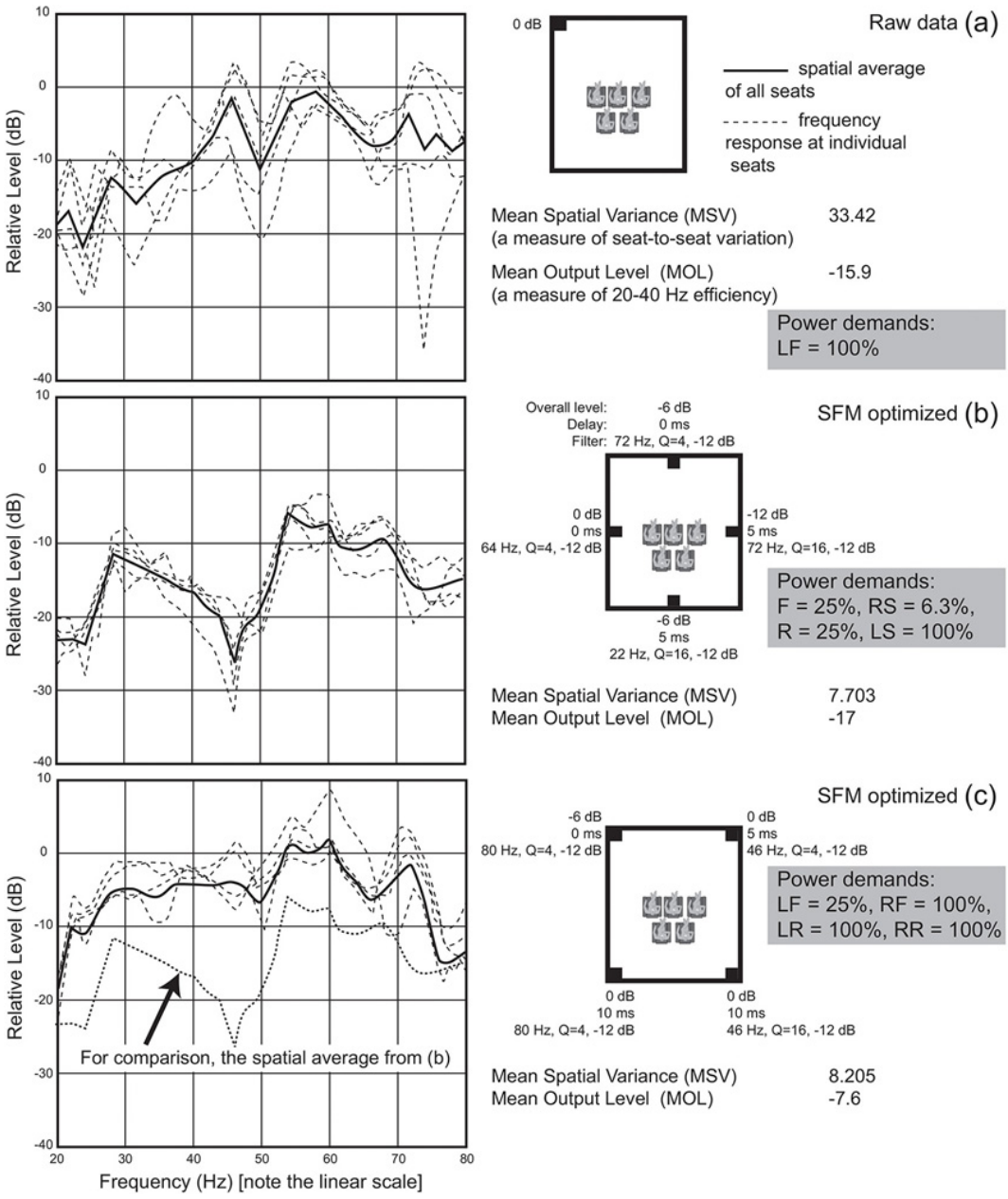


FIGURE 14.26 *Measurements made in the example room used in several earlier figures. Data from Welti and Devantier (2006): (a) their figure 20, (b) their figure 24 and (c) their figure 25. Figure 8.21 in Toole (2018).* ⏎

Again, the requirement for these strategies is a rectangular room; geometrical or structural irregularities would introduce errors. The best performance can be expected when four front-wall and four rear-wall

subwoofers are arranged to be 25% of the way from floor, ceiling and side walls, thereby attenuating the width and height modes, leaving only the length modes to be dealt with. These identical subwoofers would be mounted in or on the walls. Any of these methods benefit from measurements in the room to get exact delays for wave cancellation.

Even though half of the subwoofers are simply cancelling the other half, this method does not seem to have lowered efficiency much, since it minimizes both peaks *and dips*. Another advantage is that the specific location of seating area is not critical, since you theoretically only have a plane wave—at least at the lowest frequencies. A disadvantage is the inconvenient subwoofer locations required. See OR Chapter 14 for more simulation results comparing DBA/CABS to other methods.

14.5.8 Manipulate Modes Using Multiple Subwoofers and Computer Optimized Signal Processing—All Rooms

There are several reasons why a room may not be optimum for the solutions discussed in the previous sections:

- The room is not rectangular.
- The room is basically rectangular, but
  - There is an opening to another space.
  - There is acoustical asymmetry (e.g. one or more walls with distinctive construction, including brick, stone, expanses of glass).
  - One or more walls is not flat: a large fireplace protrusion, alcoves, etc. In custom theaters the screen wall is often a culprit; in addition to the screen, there is often custom cabinetry to house loudspeakers and equipment. In highly ornamented rooms, surfaces may be broken up with fake columns, a bar, etc.
  - Massive leather seats, arranged in rows on a staged floor, can hardly be ignored.

- Desirable subwoofer locations are not all available, or practical to use.
- Listener locations are not within the desirable areas.

In practice, such rooms are very common. What then? The pieces of the puzzle exist, and those with the time and good instincts can, by trial and error, find combinations of location and settings that provide improved listening situations in "difficult" rooms. But what is needed is assistance from ubiquitous computers to reduce the labor content in finding the right combinations of settings. While we are at it we could add more to the wish list, including allowing the use of subwoofers that are not required to be identical.

The basis of these solutions is real acoustical measurements—full complex data: amplitude and phase or impulse response—between each subwoofer location and each listener location. Then some form of signal processing is used to modify the signal sent to each subwoofer so that seat-to-seat variations are minimized. Welti and Devantier (2006) discuss some of the options. Harman's Sound Field Management (SFM) will be discussed here.

Other related methods are:

- Multi Sub Optimizer (freeware: www.andyc.diy-audio-engineering.org/mso/html/index.html)
- Trinnov Waveforming (www.trinnov.com/en/technologies/#WaveForming)
- Dirac Live Bass Control (www.dirac.com/live/bass-control/)

In SFM, measurement data are processed by an algorithm that combines the sound from all active subwoofers at each listening position (superposition). The algorithm allows for signal manipulation in the path to each subwoofer —gain, delay, and one EQ filter—with workable values pre-chosen for the parameters, and limits applied to the permitted ranges of the manipulation. Then, an optimization program is run, systematically varying the

parameters of the signal processing and monitoring the predicted frequency response at the listening positions, with the objective of minimizing the seat-to-seat variations. It is a brute-force trial and error system with some added enhancements.

An often-forgotten feature of SFM, that also seems to be missing from many other solutions, is that multiple *potential* subwoofer locations can be included in the optimization. If you have four subs and six potential locations, that means there are 15 possible ways to place those subwoofers. SFM finds the best configurations as well as DSP best settings, among up to tens of millions of possible settings. Very powerful.

To summarize, Sound Field Management has the ability to assist in the selection of optimal subwoofer positions and modify signals fed to individual subwoofers with respect to:

- Signal level (gain)
- Delay
- One parametric filter, with values of center frequency, Q and attenuation (no gain)

The optimization process starts with the operator selecting some number of potential subwoofer locations and some number of listening locations. Obviously some subwoofer locations will be better than others, as was found in the rectangular room solutions, but there is no restriction on the subwoofer locations that can be explored. The algorithm will simply do the best it can with the circumstances it is presented.

If the same subwoofer is used for the initial measurements, and it is found after optimization that some locations are not driven to full output (the gain is reduced by 6 or 12 dB, for example) this means that significantly smaller, lower power, subwoofers can potentially be used at those locations, with the only difference being lower power. A 6 dB reduction is 25% of the 0 dB power, and a 12 dB reduction is 6.3%, so

much smaller, less expensive subwoofers can be employed if desired, however they must have similar bass extension to be good substitutes.

Figure 14.26 is a good example of the impact of both the choice of subwoofer locations and SFM on seat-to-seat sound consistency and on the power required to deliver a given sound level at low frequencies. Figure 14.26(a) shows the reference condition in the same room used in computer simulations in Figure 14.19. These are real measurements in the room. The MSV is high. Figure 14.26(b) shows that employing the four mid-wall subwoofer arrangement reveals the sound level reductions attributable to the loss of the axial modes 2,0,0 and 0,2,0 (Figure 14.19(c)). The MSV, though, is dramatically reduced with SFM implemented. Substantial equalization would be needed to flatten and smooth the frequency response. In Figure 14.26(c) the higher sound output and smoother frequency response of the four-corner subwoofer arrangement is seen. MSV is at 8.2 compared to 33.42.

Focusing on system efficiency, Figure 14.26(c) shows the average room curve for configuration (b). If one wanted to roughly match the sound output from both configurations, the gain of the four-mid-wall arrangement would need to be turned up by about 9.4 dB, the difference between the MOL values of (b) and (c): −17 dB and −7.6 dB. This corresponds to a power difference of 8.7× for the 20–40 Hz frequency range over which the MOL is calculated. In practical terms, if the subwoofers used in (c) used a total of 100 watts to generate an entertaining level of low bass, condition (b) would require 870 watts. The MSV is marginally better in (b), but not enough to justify the additional power demand.

The finishing touch on all these systems is to apply global equalization to achieve a smooth, flattish frequency response. In the case of (c), the preferred arrangement, it is a matter of attenuating 4 dB peaks at 55 and 60 Hz, and a 3 dB peak at 73 Hz. The deficiency near 80 Hz would be handled in combination with the high-pass filtering for the satellite loudspeakers in the bass-managed system.

Keep in mind that this room serves as a reasonably typical example, but the results may not be directly applicable to all rooms due to variations in

room size, configuration, and acoustic properties.

The examples shown so far have been from installations in the listening rooms at Harman. Figure 14.27 is of Floyd's entertainment room constructed in 2002. It was the motivation for Welti and Devantier, his colleagues in the Harman research group, to go beyond the rectangular room multi-subwoofer solutions and solve the problem in an "acoustically inconvenient" non-rectangular room.

The magnitude of the problem can be seen in Figure 14.27(a). With the original two front subwoofers the room curves at the six seats were very disorderly, ranging 15 to 20 dB above and below the estimated average 0 dB level. Confirming audible evidence was abundant; the bass boomed and changed dramatically from seat to seat. Equalization could have improved things for one listener, but nobody else. Massive bass traps were not an option in this rather attractive room, with paintings, sculptures and fine cabinetry on display.

Adding another pair of compact, powered subwoofers and performing SFM optimization resulted in Figure 14.27(b). Not only were the seat-to-seat variations greatly reduced, but the need for global equalization was all but eliminated. Overall sound levels were significantly elevated by about 10 dB. And this was achieved with *less* total power from the subwoofers. Three of the subwoofers operated at substantially reduced levels, and with planning they could have been scaled down to considerably smaller products. How did it sound? Superb. The room, in acoustical terms, essentially disappears when "play" is pressed—no booms, only tight, deep, clean bass. Again, SFM optimization resulted in what appears from all perspectives to be desirable performance in what seemed at the outset to be a hostile situation, and all with minimal visual and physical intrusion into the living space.

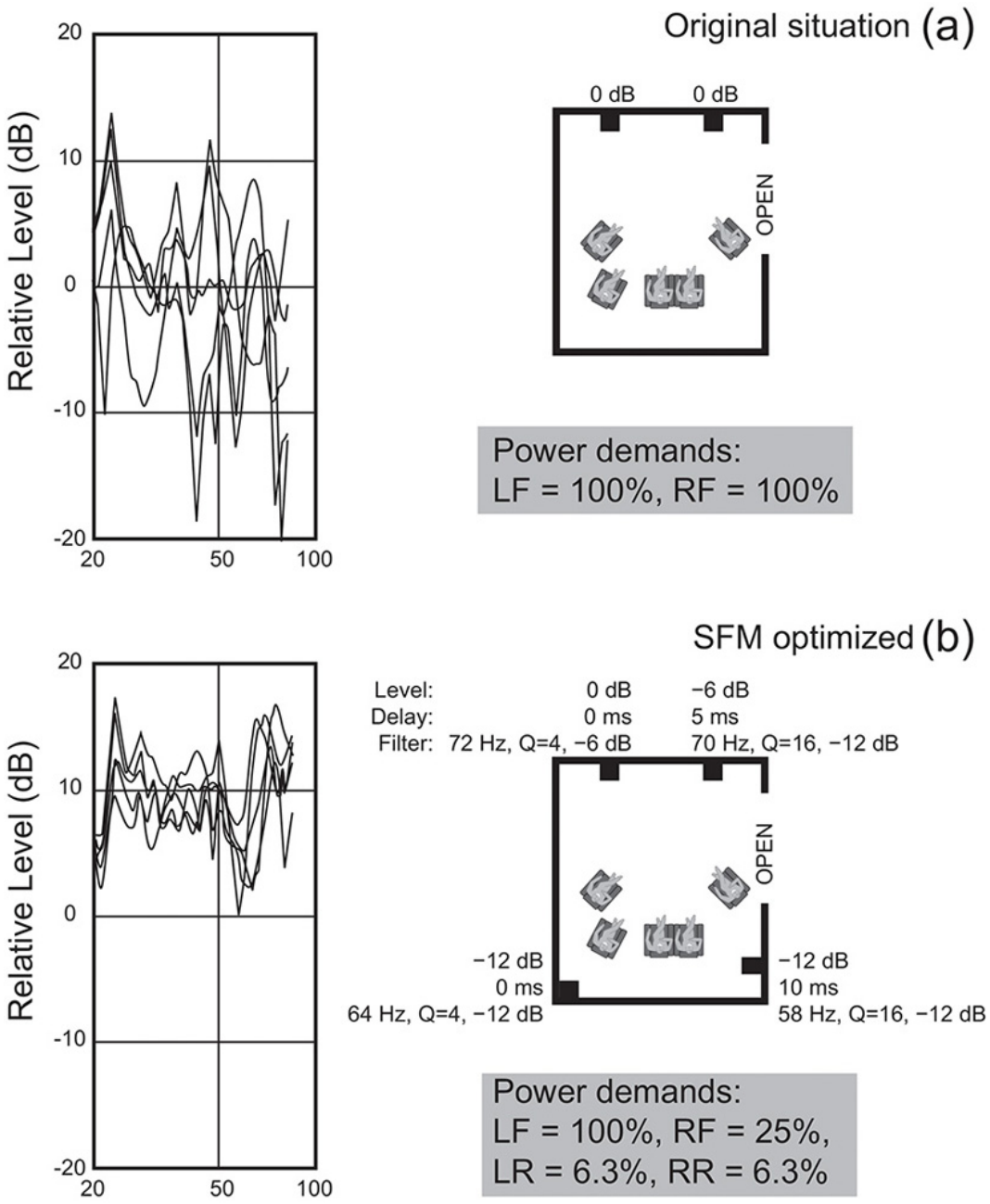
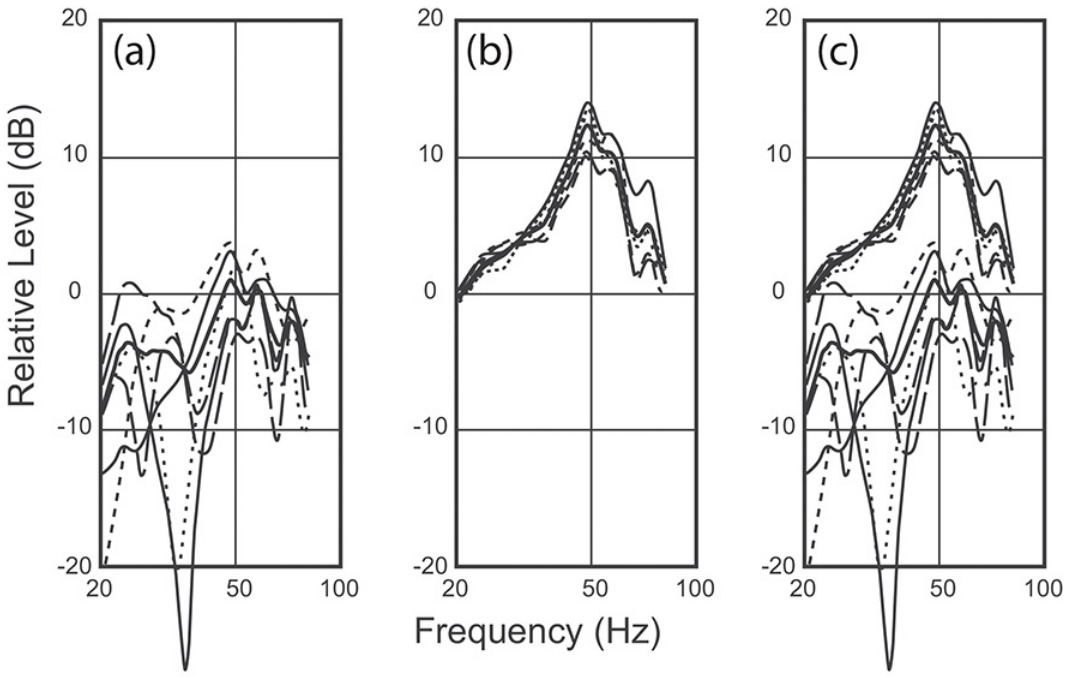


FIGURE 14.27 *Measurements made in Floyd's entertainment/family room. Subwoofers were located as space and visual considerations permitted, not by any acoustical rules. Listeners were arranged to allow verbal and visual communication with and without video. The room is 22 × 20 ft (6.7 × 6.1 m) with a ceiling that slopes from 8 to 12 ft (2.4 to 3.6 m). Data from Welti and Devantier (2003), Toole (2018, figure 8.22).* ⏎

An interesting side story: The seat closest to the lower-left corner subwoofer suffered from seriously excessive low bass in the original two-subwoofer configuration. In fact, there was a tendency to localize bass to that rear corner of the room, partly because there was so much concentrated energy there that small noises were emitted from vibrating structural elements and windows. There was no loudspeaker at that location. Adding the rear subwoofers eliminated the problem, and the subwoofers went unnoticed from a localization perspective.

Different rooms present different circumstances. Figure 14.28 shows measurements in a customer's home theater installation. It is another persuasive illustration of what can be achieved by combining multiple subwoofers with digital processing and optimization. The large seat-to-seat variations have substantially disappeared, leaving an opportunity for uncomplicated global equalization—one medium-Q parametric filter at 50 Hz—to flatten the frequency response for all listeners. The peak will be reduced by about 7 dB, meaning that in that frequency range the subwoofers will be radiating about 1/5 of their original output power. The time domain will then also be well behaved. This example employed no bass traps.

The modal control in any of these methods is achieved only on sound reproduced through the engineered system, but that is the objective. The fact that the processing delivers good sound to several locations in the room means that the sound field between and around those locations tends also to be improved. One can play a loop of kick drum and bass guitar and walk around the chosen seating area with little variation in bass timbre or tightness.



(a) One corner subwoofer, measured at six seats
(b) Four corner subwoofers with SFM, measured at six seats
(c) (a) and (b) combined

FIGURE 14.28 *In a home theater (a) shows steady-state room curves at six seats using a single subwoofer. (b) shows the measurements with four subwoofers with levels, delays and one parametric filter for each subwoofer selected by a computer optimization program—Harman's Sound Field Management. (c) shows a direct comparison. No global equalization in these measurements. Toole (2018, figure 8.23).* ⏎

The only unfortunate fact is that, at time of publication, SFM is only available in JBL Synthesis custom installation products. The good news, though, is that using trial and error, good measurements and patience, satisfying results can be achieved by do-it-yourself enthusiasts. Figure 14.29 shows the results obtained in a difficult L-shaped room using five subwoofers and manually adjusted digital processing (DellaSala, 2016). In (b) the bass is very uniform, gently rising and, although not shown, it extends well into subsonic frequencies. Different rooms will obviously require different solutions.

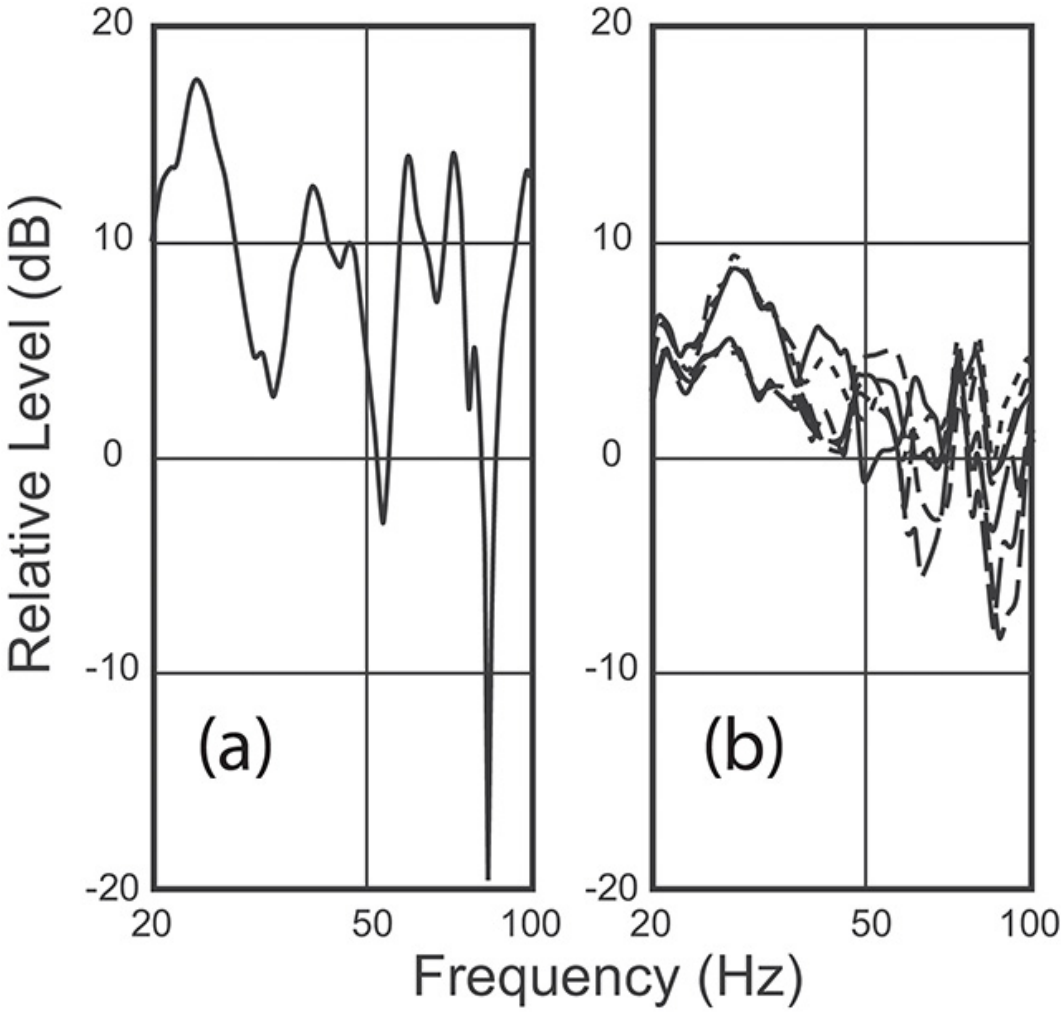


FIGURE 14.29 *(a) The steady-state room curve at the prime seat using the original front loudspeakers. (b) The results at all six seats after trial-and-error placement of additional subwoofers and delay, gain and equalization adjustments using a digital processor. Sound levels in (a) and (b) cannot be compared. Data courtesy Gene DellaSala, Audioholics.com. Toole (2018, figure 8.25).* ⏎

14.5.9 Final Equalization at Low Frequencies: Global Correction of System Response Without Modal Modification

In this section, we explore the electronic manipulation of an electro-acoustic phenomenon. This is made possible by the fact that low-frequency

room resonances behave as minimum-phase systems, allowing for minimum-phase equalization to reduce time-domain ringing. It's important to note that this does not alter the acoustic resonance itself; the correction applies only to *reproduced* sound. Equalization is most effective at the listening position where measurements are taken. For other listeners in the room, the experience may vary significantly, with seat-to-seat differences remaining substantial, just as they were prior to equalization. However, if seat-to-seat variations can be minimized, broader area equalization may be achievable.

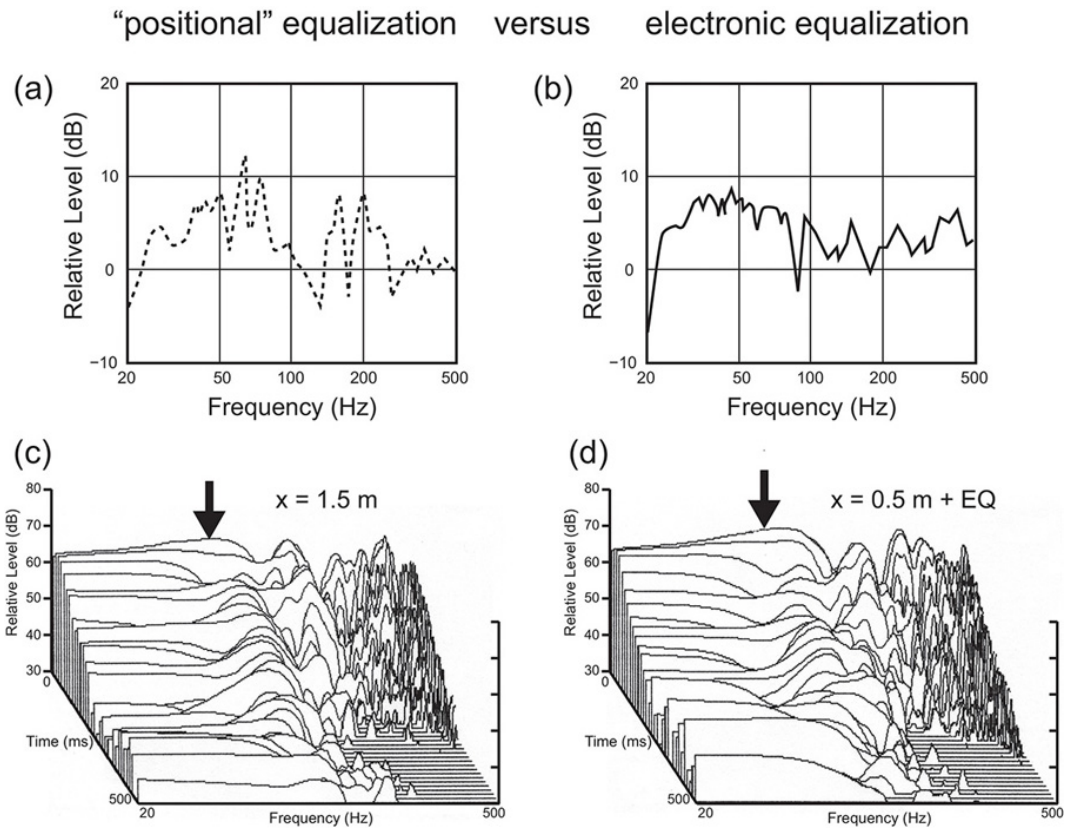


FIGURE 14.30 *A comparison of steady-state frequency responses (a) and (b) and waterfall diagrams (c) and (d) for two methods of addressing a problematic room resonance. Toole (2018, figure 8.11).* ⏎

Figure 14.30 shows data comparing approach (2, positional adjustment) with the equalization method in Floyd's large listening room. On the left,

(a) shows the frequency response and (c) the waterfall for "positional" equalization, with the listener at 1.5 m from the back wall (from Figure 14.15(b) and (d)). Figure 14.30(b) and (d) on the right show the comparable results for the listener moved back to 0.5 m from the wall and matched parametric equalization engaged. The main point of the figure is to show how very similar the waterfall plots look. Most of the small differences in both the frequency responses and the waterfalls are the result of making the measurements at different locations, with the consequent different interactions with other standing waves in the room, none of which were as consequential as the "monster" mode at 42 Hz (arrow).

Looking similar is one thing, but how did they sound? Over several months, colleagues, audio journalists and interested social visitors were subjected to simple A vs. B comparisons of the two conditions. The overwhelming conclusion was that they sounded remarkably similar in every respect and very much better than the original condition. The most dramatic demonstration sound was a well-recorded kick drum that, before treatment, was amusingly fat and flabby and, after treatment, became an abrupt "slam"—as it should be. It was audibly obvious that the time-domain problem had been repaired.

Equalization had the huge advantage of allowing the listener to sit in a sensible location. Acoustically, there were advantages too. With up to a 14 dB amplitude reduction around 42 Hz, the woofers no longer had to work so hard, distortion was lower, and they could play louder. There was also *much* less energy everywhere in the room at 42 Hz.

Was the mode eliminated? No, but audible evidence of it was removed when the sound system was operating. It only works for sounds radiated by the loudspeakers, not from a bass guitar that someone might pick up and play. It also only worked for one seat, and one on each side on a sofa, parallel to the length mode. The room acoustics were unchanged.

If this solution is chosen it is important to know what kind of equalizer to use. Figure 14.31 illustrates what happens when measurements on a subwoofer in a room lack the resolution to reveal the problem in sufficient detail, and when the equalizer lacks the resolution to address the resonance. It is evident that high-resolution (1/20-octave in this case) measurements were necessary to reveal the true nature of the resonance and that matching the Q of the resonance with the parametric equalizer had a significant beneficial effect on the reduction of time-domain ringing. Excessive smoothing of room curves disguises the true nature of the resonance and prevents a maximally effective equalization. For similar reasons, a graphic-type equalizer with fixed filters, is not effective.

14.5.10 Revisiting Room Resonances in Time and Space

Figure 14.32 looks at the current topic from a slightly different perspective, illustrating how standing waves associated with room resonances are revealed in multiple measurements along the specific trajectory shown in Figure 14.32(a). Figure 14.32(b) shows results *using only subwoofer* 1. It can be seen that:

- Length mode 1,0,0 is weakly energized because the woofer is in the null of the mode, and is attenuated further because the microphone trajectory runs along the null line across the width of the room.
- Width mode 0,1,0 is strongly energized and shows decaying amplitude as the microphone approaches the null at the halfway point (see also bottom of (a)).
- Tangential mode 1,1,0 is not very strong and also shows declining amplitude as the center of the room is approached.
- Length mode 2,0,0 is strongly energized and the amplitude remains steady across the width of the room.
- Width mode 0,2,0 and tangential mode 1,2,0 are very close (56.5 and 54 Hz) and in the modal interaction one can see a slight frequency shift

in the dominant peak as the microphone is moved to the center of the room.

- Height mode 0,0,1 is poorly represented because the microphone was close to the vertical null.
- Length mode 3,0,0 is essentially absent because both the woofer and the microphone trajectory are in the null of the mode.
- Tangential mode 2,2,0 is strongly energized and shows an untidy set of curves as the microphone is moved through a null and then back to a maximum, as can be seen in Figure 14.20.

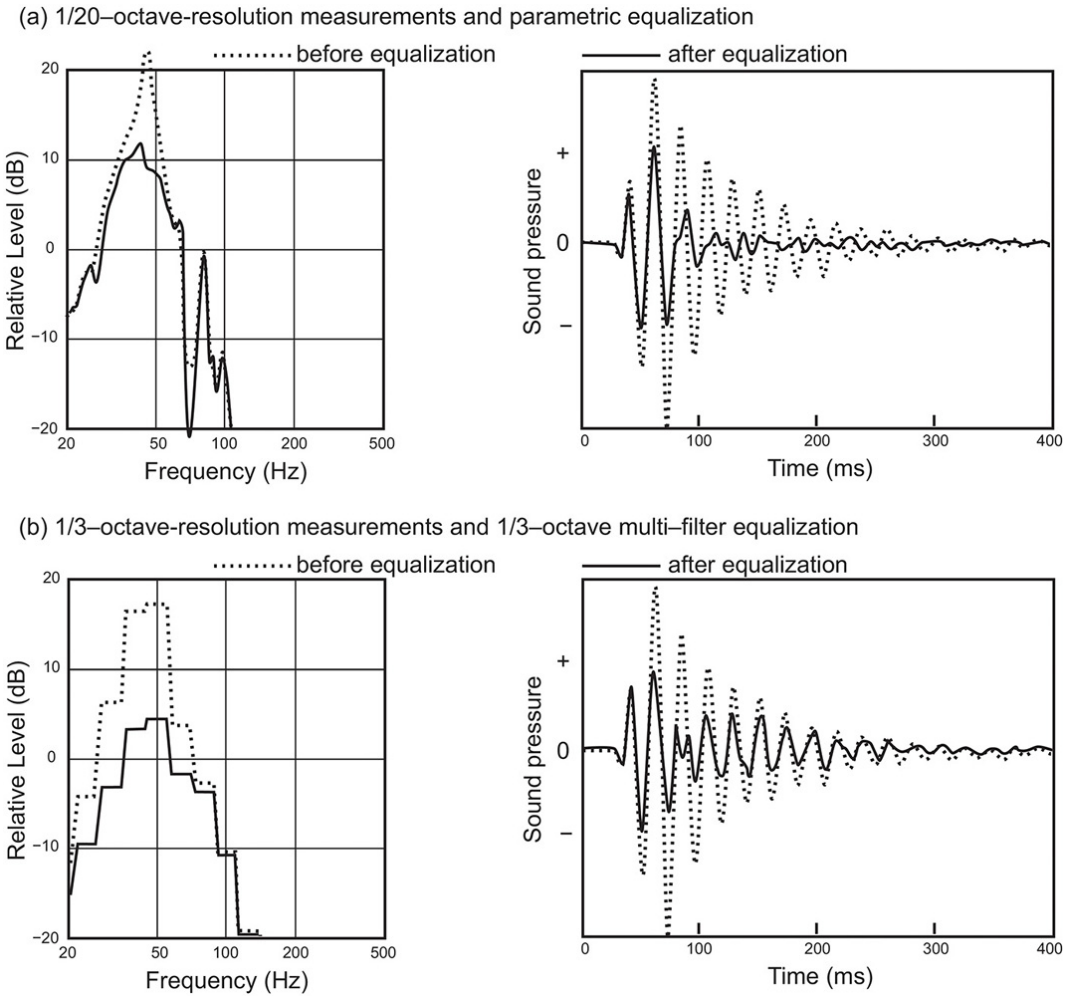
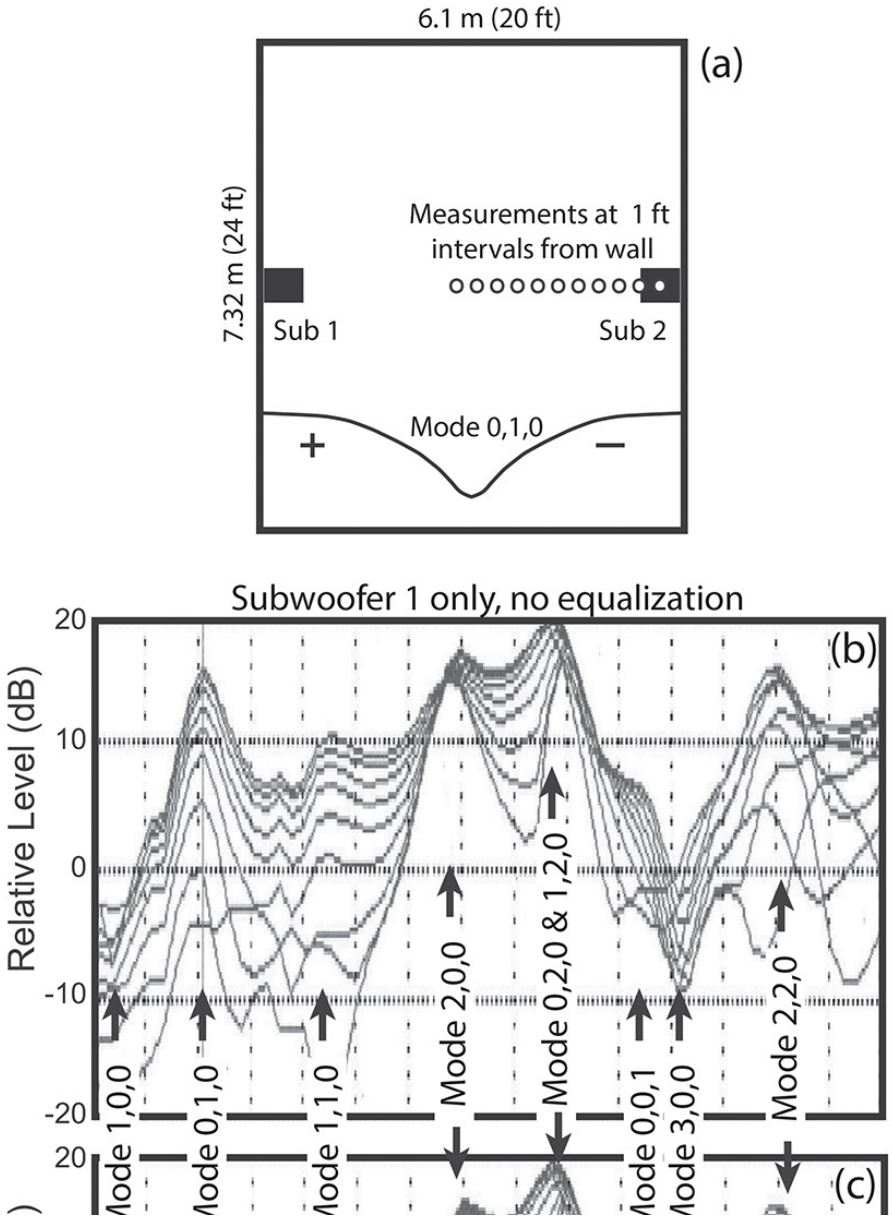


FIGURE 14.31 *Before and after measurements in the frequency and time domains for two different measurement and equalization*

methods. This is figure 8.12 in Toole (2018). ⏎

This brief analysis is a good example of just how predictable modal events are in a known rectangular room when the woofer and microphone locations are well defined. This underlines the huge advantage of rectangular rooms—the low-frequency sound fields can be substantially predicted in advance and measurements interpreted after the fact. Knowing which modes cause problems allow solutions to be focused. Obviously, only high-resolution measurements allow for this depth of analysis.



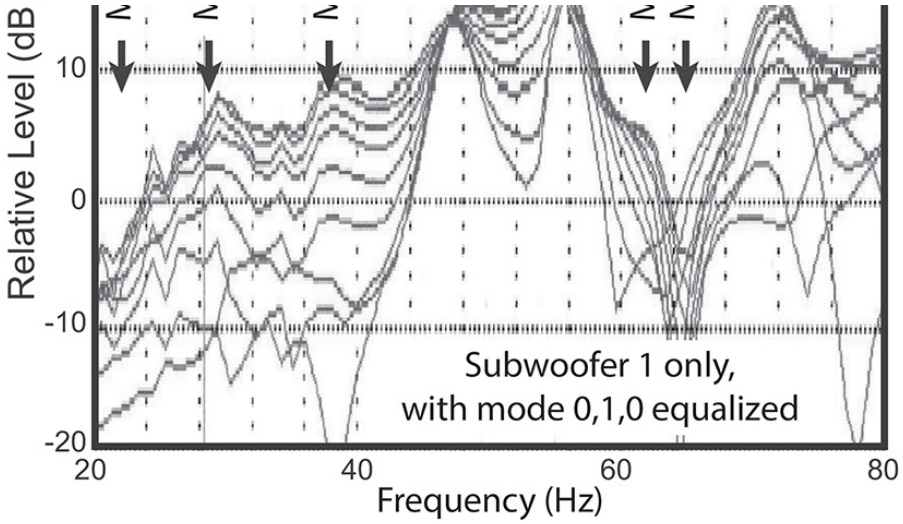


FIGURE 14.32 *(a) shows the room layout with an illustration of the width mode being addressed. Unlike the stylized mode illustrations in Figure 14.17, this is a sketch of the sound level along the measurement axis using peak values extracted from the curves in (b). NOTE: Only subwoofer 1 was used for the measurements in this figure. In that location mode 0,1,0 is maximally energized. (b) shows measurements made at the 1-ft intervals starting at the side wall and moving towards the center. The resonant peak is evident in all of the curves, as is the declining sound level as the microphone is moved towards the null in the center of the room. (c) shows the attenuation of the peak by parametric equalization, and confirms that the beneficial effect is seen in all of the measurements. There is no change in the spatial distribution.* ⏎

The experiment involved using a parametric filter to attenuate mode 0,1,0. The filter used was not a perfect match for the mode—it had a slightly lower Q—so the attenuation is less than it might have been. Perhaps this makes it even more of a real-world example. It is first worth examining the plot of sound pressure vs. distance at the bottom of (a). This is derived from the peak sound levels of mode 0,1,0 as a function of distance from the wall as shown in (b). The notable feature is the flatness of the pressure

peaks compared to the narrower interference dip. The lesson is that resonant peaks are bothersome because they are resonances, with accentuated amplitude and ringing, and also because the peaks occupy most of the floor area of the room. More listeners are annoyed by peaks than dips for this reason and also because the presence of excessive sound is more annoying than an absence—see discussion of peak and dip audibility in Sections 9.6.1 and 9.6.2. The measurements in Figure 14.32(c) indicate that the effect of the mode is attenuated at all microphone locations, which is to be expected, and also that none of the other modes is significantly affected. Substantial bass energy has been removed; perhaps more than is optimum for good listening.

If one were to equalize this system it would probably be reasonable to start by aiming for the 10 dB horizontal line as a target, attenuating mode (0,1,0) by about 6 dB, and introducing other parametric filters to attenuate modes (2,0,0), (0,2,0) and (1,2,0) combined, and (2,2,0) to the 10 dB line. Attempts to fill the dips associated with the (0,0,1) and (3,0,0) modal nulls are not advised. This is why mode cancelling and manipulation techniques are so advantageous—they reduce peaks and elevate nulls while using less energy. After the "high resolution" equalization is complete, broadband spectrum adjustments can be made to bring the result into agreement with the desired overall room curve target, as discussed in Sections 12.3.3 to 12.3.5. Automated low-frequency equalization algorithms should be capable of doing both operations simultaneously, but some are overly zealous in trying to achieve a smooth target curve. This measurement also highlighted what we already knew: global equalization does not affect seat-to-seat variation!

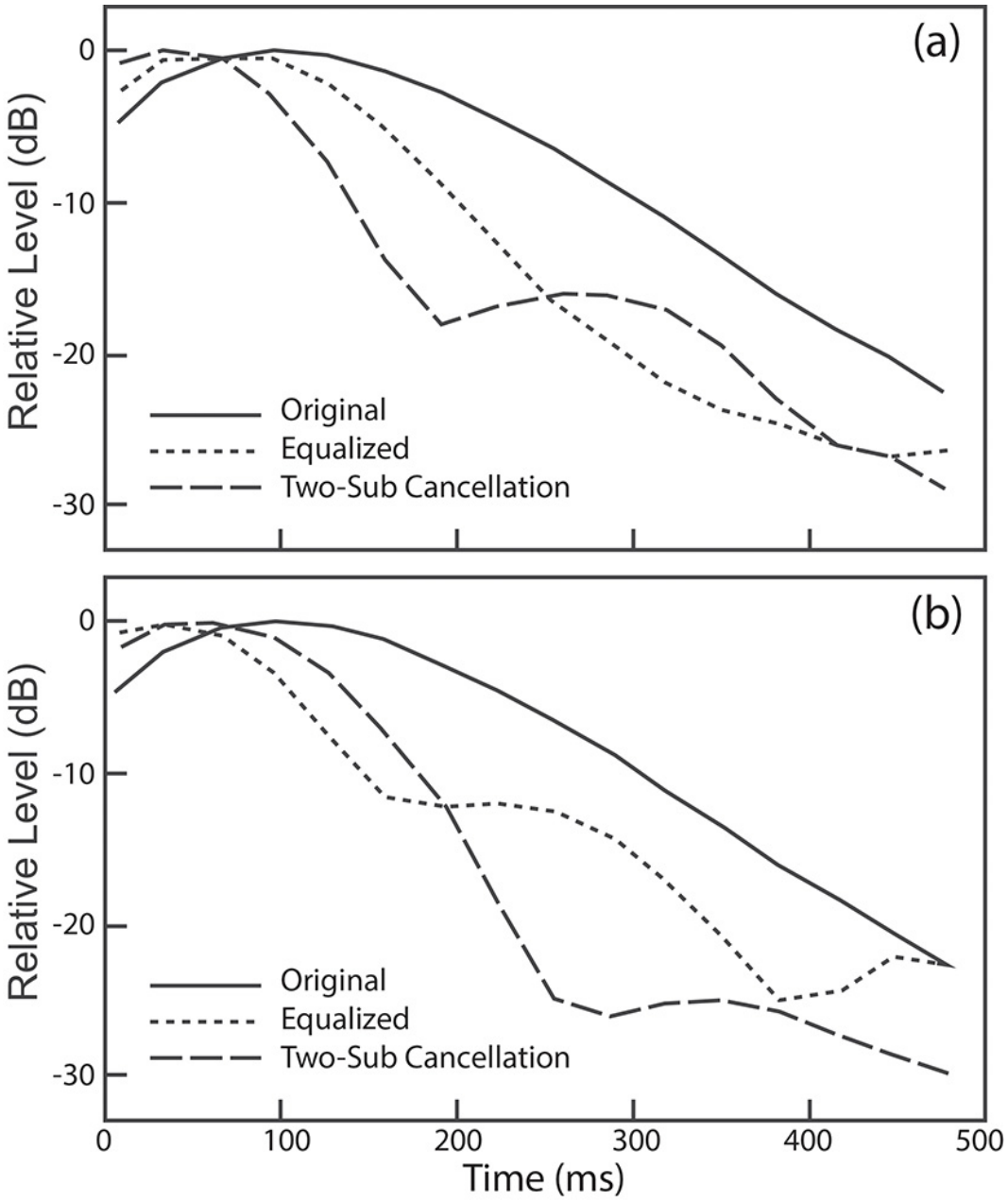


FIGURE 14.33 *Energy-time curves measured at (a) 1 foot from the wall in the floor plan shown in Figure 14.32, and (b) at 10 feet from the wall, the mid-point of the room width. The signal was filtered at the center frequency of mode 0,1,0, so what is shown reveals the time course of sound decaying at the modal frequency in three conditions: the original modal decay, the decay after the mode was attenuated by parametric equalization (as shown in Figure 14.32), and after the mode was attenuated by*

employing subwoofers 1 and 2 in a "mode cancelling" configuration, driven with identical signals. ⏎

Figure 14.33 is especially interesting as it shows the time-domain performance of two methods of attenuating a room resonance: equalization of one sub and mode cancelling using two subs (1 and 2 in Figure 14.32(a)). These energy-time curves illustrate the amplitude vs. time events at the frequency of resonance (0,1,0), seen above. The data are not perfect because the filter used to isolate these events in the frequency domain has a decay time of its own, but the differences between the curves are valid. In this comparison the curves have been adjusted so that they all exhibit the same peak sound levels. In reality the equalized and mode-cancelled curves would start at lower levels, thereby adding further assurance of reduced audibility. Figure 14.32(b) and (c) show a 9 dB reduction in amplitude due to the equalization, to which the increased decay rate is an added improvement. It is interesting to see the buildup of the resonant energy in the "original" situation, as was seen in the tone-burst waveforms of Figure 5.4. All of the bass energy at this frequency must be communicated through the resonance, and it has a build-up time and a decay time. Both the equalization and cancellation treatments reduce the build-up interval and the ringing, but they behave differently in their interactions with the prevailing sound field. In both cases it is not difficult to hear the improvement in bass uniformity and temporal definition.

Looking back at Figure 14.30(d) for equalization and Figure 14.18(e) for cancellation, one can see the waterfalls drop quickly in the early time interval moving forward in the illustrations. They then stabilize to a lower decay rate. This is also seen here, especially in the mode cancellation situation. The initial drop in all cases is about 15 dB or more which is perceptually very significant. Recall that at these low frequencies perceived loudness halves for each 5 dB or so of amplitude reduction (Section 9.4). When listening to music or movies, the resonance is perceptually insignificant.

14.5.11 Do We Hear the Spectral Bump, the Temporal Ringing or Both?

The preceding sections have enumerated and explained several ways to attenuate bothersome room modes, ways to manage which are active and which are not, and electronically elegant ways to deliver similarly good bass to several listeners. Because low-frequency room resonances generally behave as minimum-phase systems individual modes can be addressed with parametric-filter equalization to reduce the magnitude of the spectral bump associated with a room resonance. If there is a bump, there will be ringing. If the bump is attenuated by equalization, or, in fact any of the alternative schemes, so is the ringing. Waterfall diagrams and impulse responses routinely confirm the improvements in time-domain behavior.

In many cases, as shown in the preceding examples, the problems and the improvements are not difficult to measure or to hear. As is seen in Figure 14.33, a resonance, in this case a medium-Q resonance, does not respond to an input instantaneously. There is a build-up interval as well as a decay interval, the former occurring while a signal is present (see also Figure 5.4). A kick drum may appear to be a sharp transient, but the impulsive "kick" is at frequencies much higher than the fundamental resonance "boom" which can last 200 ms or so. What sounds "tight" is actually a drum-head/volume resonance superimposed on room resonances. During that time the amplitude at the room resonance frequency increases before the ringing decay begins; the higher the Q, the longer the build-up interval, and the longer the ringing. So, the perception of such an event is not a simple one. The notion of "fast" subwoofers is misguided, in that the time-domain performance of minimum-phase systems is predictable from the amplitude response. With an 80 Hz bass management crossover frequency, all subwoofers are equally "fast". So, what we confront in the real world is essentially "fast" subwoofers communicating through "slow" rooms.

Several studies have examined what happens in the time and frequency domains when tailored filters address specific modes as distinct from what happens with "typical" room equalization, where for historic reasons the

frequency resolution is reduced from what is easily achieved today, to 1/3-octave smoothed curves. Obviously, with less precise data, the equalization is not as well matched to the resonance, and the ringing is less controlled (see Figure 14.31).

But how much precision is necessary? In real rooms several modes can be simultaneously active and stimulated by ever-changing music containing long and short tonal components and various forms of transient stimulation. There can be subtle things like two adjacent modes that share energy, one decaying faster than the other and effectively shifting the pitch of the resonance (Figure 14.18(c), (d) and (e)). If they are of similar Q, they will beat as they decay. A mode can have high Q, long ringing, but be at a level significantly below the overall spectral level, as in Figure 14.18(e) and others. These are all situations difficult to subjectively examine. So, investigations that have been done tend to be conducted under very controlled circumstances, most involving computer simulations of both the problems and the solutions.

The following experiments involve comparisons of various forms of equalization, but it is important to note that they address audible performance only for a single listener. Just one test employed a multi-subwoofer mode-manipulation scheme and it turned out to be more effective than necessary from a mode attenuation perspective, but of course had the unique virtue of creating multiple seats where similarly good bass can be heard.

Toole and Olive (1988) concluded that at frequencies above 200 Hz the threshold of detection of resonances decreased about 3 dB for each doubling of the Q value. This means that the higher the Q, and therefore the longer the ringing, the less detectable the resonance. This is not an indication that ringing per se is not a problem, rather it most likely relates to the probability of a narrow band resonance being energized by typical music. Consequently, at these frequencies the duration of ringing is not a reliable indicator of audibility—see Section 5.3.1.

Those investigations were in the frequency range above the transition frequency, where reflected sounds were an important factor. Reverberation

—reflected sounds—provide masking for high-Q ringing, and perceptual amplification for medium- and low-Q resonances by providing repetitions of the events. Reverberation time measurements at these frequencies are dominated by the decay of reflected sounds: repetitions of the direct sound. At bass frequencies the ringing from strong resonances is likely to be the dominant factor in RT. This is a continuation of the direct sound, not reflected repetitions of it coming from different directions. Binaural hearing is not very effective at low frequencies, so measured RT can have different perceptual consequences at different frequencies. This is often ignored in discussions.

Olive et al. (1997) elaborated the study to include lower frequencies, finding again that increasing Q yielded increasingly high detection thresholds as revealed by the amplitude peak in a room curve. Decreasing frequency also yielded increasing thresholds except for the lowest Q resonances. Q = 1 resonances were detected at low threshold levels whether they were resonances (peaks) or anti-resonances (dips) but it must be noted that such low-Q, wide-bandwidth, resonances embrace most of the frequency range of a typical subwoofer, so this result is hardly surprising. It really is changing the overall spectral balance in an unsubtle way. Here it was found that impulsive signals increased the audibility of high-Q resonances, suggesting that ringing was itself a contributor to audibility. None of this is surprising, but more detailed information would be useful. Fortunately, other investigators have added insights, but with so many variables simultaneously at play it is difficult to be absolutely definitive.

Antsalo et al. (2003) compared the audibility of simplistic 1/3-octave smoothed room curve equalization to that equalization combined with modal equalization that addressed specific room modes with higher precision. The waterfall diagrams show that the synthetic modes were usefully addressed by the modal equalization, reducing both the initial amplitudes and the decay times of the ringing. In simulations over headphones, listeners compared the original room with the room curve smoothed by the magnitude equalization, and a combination of that curve and modal equalization achieved by two different methods. The result was

that listeners reported little or no improvement over the simple smoothed magnitude equalization, which alone could only reduce the spectral prominence of the resonances, without addressing the decay rate. The inference is that listeners responded to the reduction of the "bump" rather than the reduction of the decay rate, and did so even with 1/3-octave equalization. The tests relate to a single listening location.

Karjalainen et al. (2004), in work shared with the previous authors, varied the duration of ringing in modes that had been amplitude equalized to within ±3 dB of the 1/3-octave smoothed response that left the modal ringing essentially intact. The room had a reverberation time of 0.28 s at mid frequencies, rising gradually to 0.5 s at 50 Hz. Test signals included speech, 0.5 s noise bursts, drum hits, and rock music. They found that the detection thresholds of ringing decay times were above the RT of the room, rising gradually from about 0.5 s at 800 Hz to about 0.7 s at 100 Hz. Incredibly, below 100 Hz the listeners became unresponsive to modal decay times up to almost 2 s. Noting that the amplitude responses had been equalized, they concluded that: "Without such equalization the differences can be noticed much more easily, which indicates the obvious fact that the frequency-domain properties of resonances are prominent over the time domain properties." This is in agreement with the Toole and Olive finding for resonances above 200 Hz. In any event, the conclusion was that modal ringing at frequencies below about 100 Hz is not an issue—but there is more.

Avis et al. (2007) examined the audibility of a room resonance as a function of Q, without specific control in modeling or in listening test programs that favors either decay time or spectral variance. Q is translatable into a corresponding decay time, so there is continuity of intent. A summary result is that "reductions to the Q factor below a value of about 16 will be subjectively imperceptible" which translates into calculated modal decay times of about 1 s at 32 Hz, 0.7 s at 50 Hz and 0.28 s for 125 Hz. In the context of normal listening spaces the high-frequency value will likely be below the measured RT, so there is the matter of masking of the resonant ringing by room reflections. The authors addressed RT as a variable, but the

data applied only to frequencies above 250 Hz, leaving the key question about simultaneous masking at the resonance frequencies unanswered. Unless the room is inherently very dead, as happens in a few recording control rooms, these durations of ringing may not be audible. There is also the uncertainty of the height of the resonant peak above the average spectrum level; i.e. how much does it perceptually stand out? In any event, these data add to the impression that ringing at very low frequencies is not easy to hear.

Adding further information and complexity to the discussion, Fazenda et al. (2012) subjectively and objectively compared several methods of addressing room modes. Unfortunately, they chose to degrade the measurement prior to equalization to 1/3-octave resolution. They then noted that this had detrimental effects in both the high-resolution frequency responses and in the decay rates of resonances. So, why was it done? The most useful observation was that the active multiple-subwoofer mode-control scheme CABS, discussed in Section 14.5.7, worked very well—even better than they thought was required by resonance detection criteria they had developed. This is in principle similar to SFM discussed in Section 14.5.8. So, if resonance detection at a single listening location is the only thing of value, the electronically processed mode control methods might be too good, but the additional value they bring is that multiple listeners can share the excellence, which cannot be said of other schemes.

On the practical side, they point out that improvements are possible by relocating loudspeakers and, if that is not possible, by simple magnitude equalization. In fact, simple equalization is advisable even when repositioning is possible. Obviously, it is time to abandon low-resolution (1/3-octave) measurements, even though they exist in legacy industry standards. All of the data indicate that modal equalization is superior, especially in demanding listening situations.

Fazenda et al. (2015) continued the investigation of the audibility of the time decay component of resonances. They examined resonances in isolation and in simulated room contexts, finding that with music stimuli the thresholds were 0.51 s at 63 Hz, 0.3 s at 125 Hz and 0.12 s at 250 Hz. In

practical terms this means that in the subwoofer frequency range (typically below 80 Hz) these decay times can very likely be achieved with conventional parametric equalization of problem room modes, and almost certainly with any of the multiple subwoofer active control schemes that bring the additional and important advantage of providing several listeners with similarly good bass sound.

All these studies indicate that parametric equalization based on high resolution steady-state frequency responses is a reliable method to attenuate bothersome room resonances. Waterfall diagrams alone are of little value in estimating the audible consequences of resonances. The uncertainties in what one commonly sees in the decorative waterfalls adds to the confusion (see Figure 9.11).

14.5.12 Stereo Bass: Little Ado About Even Less

Apologies to William Shakespeare. This issue relates to the fact that, in order for all of the systems described here to function fully, the bass must be monophonic below the subwoofer crossover frequency, commonly 80 Hz. Most of the bass in common program material is highly correlated or monophonic to begin with and bass-management systems are commonplace, but some have argued that it is necessary to preserve at least 2-channel playback down to some very low frequency to preserve aspects of spatial effect at very low frequencies. There is perceptual evidence of spatial/enveloping information in large concert halls, in which long wavelength sounds can arrive from many directions, reducing interaural cross correlation. Numerous experiments show that the binaural auditory system responds to interaural differences in amplitude, phase and timing down to very low frequencies, including those that occur naturally at nulls in standing waves. The perceptual basis for hearing spatial properties in low bass (stereo bass) is real, but perspective on its importance in everyday listening is unclear.

The questions are:

- Can this be communicated to listeners in small rooms?

- Is there a frequency (the subwoofer crossover frequency) below which the effect is no longer important?
- If low-frequency envelopment information is important and can be communicated, can it be done without degrading the timbral improvements delivered by multiple subwoofer schemes? Perceptual priorities might be involved.

Experimental evidence thus far has been mixed (Welti, 2004). The authors have participated in a few comparisons, carefully set up and supervised by proponents of stereo bass. Some effects were heard, mostly subtle, with the audible spatial effects attributed to frequencies above about 80 Hz. This means, in essence, that it is a "stereo *upper*-bass" issue, and the surround channels (which typically operate down to 80 Hz) are already "stereo" and placed at the sides for maximum benefit. We listened in a well-damped, demonstration listening room.

Martens et al. (2004) identify a "cutoff-frequency boundary between 50 Hz and 63 Hz", these being the center frequencies of the octave bands of noise used as signals. Recent discussions among knowledgeable people in an audio forum revealed strong opinions in favor of spatial low bass, especially in classical recordings. (https://www.audiosciencereview.com/forum/index.php?threads/bass-and-subwoofers.51589/, accessed 05/2025). Thomas Lund discusses his perspective: https://assets.swoogo.com/uploads/4646601-673c03ab61c93.pdf, accessed 05/2025). It appears that small room reflections/reverberation at low frequencies require substantial low-frequency absorption to avoid masking the desired effects, and placement of loudspeakers to the sides of the listeners maximize the effect, as Welti (2004) found. We are not aware of definitive subjective/objective experiments providing guidance, nor any clarity as to how the low-bass directional/spatial information can be reliably captured and delivered. It seems to be work in progress, but as stated earlier one hopes that it can be

accomplished without degrading the great improvements in bass sound quality and seat-to-seat consistency that multiple subs can deliver.

14.5.13 Bass Management Makes It All Possible

All the multiple subwoofer schemes discussed earlier in this chapter assume that there is a single signal for all bass frequencies, no matter how many satellite channels there are. That single channel may be subdivided into more for digital signal processing used in some more advanced multiple subwoofer solutions.

The key to these solutions working is in the bass management feature in all surround processors and many stereo processors. It combines the low bass from all channels, adds the LFE if it is present, and delivers the combined signal to a subwoofer channel output. Bass management is explained in Section 8.2.2. There has been continuing research in this area (see OR Chapter 14). In the end, any possible modest spatial effects obtained may come at a significant price.

14.6 SUMMARY AND DISCUSSION

The basic techniques discussed above are all in the scientific literature, some for decades. Commercial acoustical absorption has been around since the 1930s, when dozens of companies were manufacturing versions of resistive absorbers. The design of bass traps came later and fit the general pattern of acoustical treatment being synonymous with acoustical absorption. “Dead” acoustics became fashionable. Fortunately, we have moved on from those times. But bass in small rooms remains a challenge.

New generations of audio enthusiasts and experts need to become aware of current knowledge and learn how to use it to their advantage. Unfortunately, internet forums continue to have protracted debates over issues long settled, good measurements are rare, opinions abound, and those with business interests continue to agitate to their advantage. Our observations of the professional audio field indicate that the situation is similar.

One of the most common misconceptions is that simply adding another subwoofer or two will automatically improve things. Most often they are not identical subwoofers, they are not located where they need to be, and they may or may not be driven with the same bass-managed signal. Unpredictable results and debates are almost inevitable.

When measurements are done, it is common for some individuals to try equalizing each subwoofer separately and then to combine them. Again, disappointment is almost guaranteed because when the system is in "play" mode all subwoofers are operating simultaneously, and the result of the acoustical summation is unpredictable. The measurement that matters is the one done with all subwoofers operating simultaneously. For individually measured subwoofer outputs to sum in a predictable manner, the measurements must be of the transfer function—amplitude and phase—not the steady-state amplitude response. This is the basis of the room mode manipulation schemes described in Sections 14.5.7 and 14.5.8, in which after individual subwoofer measurements are done at the seating locations, predictions of room response can be calculated in the computer with high precision, including experimenting with delay, gain and equalization in each of the subwoofer signal paths. Acoustical superposition applies.

There is also the "spatial averaging" confusion. By measuring at several seats and averaging the result, one gets a spectrally smoothed overall impression of what the audience hears, but not of what any individual in the audience hears. Unless one employs a massive bass-trap array or a multiple-subwoofer strategy to reduce seat-to-seat variations, equalization to satisfy a spatial average may end up truly satisfying none of the listeners. It is arguably better in such cases to satisfy the prime listener, and let the rest take their chances. The latter approach is a common approach if one wishes to optimize the subwoofer-to-satellite crossovers. With digital processors one can have both "group" and "just me" equalizer settings.

So, let us summarize this section about bass by saying that with intelligent use of multiple subwoofers there is now an alternative to or an addition to bass traps. Low-frequency absorbers work and will always make things better if one chooses one of the multi-subwoofer options. But the

reality that faces most people in homes is that large passive low-frequency absorbers are difficult to accommodate. Multiple subwoofers also present complications. Nevertheless, to the extent that absorption can be installed, starting with a smaller problem makes any additional solution work better. Acoustical absorbers remove energy, multi-subwoofer solutions tend to add efficiency, so there is a trade-off to consider. The benefactors are the listeners.

REFERENCES

Antsalo, P., Karjalainen, M., Mäkivirta, A. and Välimäki (2003). "Comparison of Modal Equalizer Design Methods", *Audio Eng. Soc., 114th Convention*, Preprint 5844. ⏎

Avis, M., Fazenda, B. and Davies, W. (2007). "Thresholds of Detection for Changes to the Q Factor of Low-Frequency Modes in Listening Environments", *J. Audio Eng. Soc*., vol. 55, pp. 611–622. ⏎

Bolt, R. H. (1946). "Note on Normal Frequency Statistics for Rectangular Rooms", *J. Acoust. Soc. Am.*, vol. 18, pp. 130–133. ⏎

Celestinos, A. and Neilsen, S. (2008). "Controlled Acoustic Bass System (CABS), a Method to Achieve Uniform Sound Field Distribution as Low Frequencies in Rectangular Rooms", *J. Audio Eng. Soc.*, vol. 56, pp. 915–931. ⏎

DellaSala, G. (2016). "Bass Optimization for Home Theater with Multi-Sub + mDSP". www.audioholics.com/home-theater-calibration/bass-optimization-for-home-theater, March 3, 2016. ⏎

Fazenda, B., Stephenson, M. and Goldberg, A. (2015). "Perceptual Thresholds for the Effects of Room Modes as a Function of Modal Decay", *J. Audio Eng. Soc*., vol. 137, pp. 1088–1098. ⏎

Fazenda, B., Wankling, M., Hargreaves, J., Elmer, L. and Hirst, J. (2012). "Subjective Preference of Modal Control Methods in Listening Rooms", *J. Audio Eng. Soc*., vol. 60, pp. 338–349. ⏎

Geddes, E. R. (1982). "An Analysis of the Low Frequency Sound Field in Non-rectangular Enclosures Using the Finite Element Method", PhD

thesis, Pennsylvania State University. ⏎

Geddes, E. R. (2005). “Audio Acoustics in Small Rooms”, a PowerPoint presentation available at www.gedlee.com ⏎

Goertz, A., Wolff, M. and Naumann, L. (2001). “Optimization of Sound Reproduction in Listening Rooms for Surround Sound Loudspeaker Setups”, White paper for Klein+Hummel. English translation at: http://forums.klipsch.com/forums/storage/8/1485263/tmt2002_eng.pdf ⏎

Karjalainen, M., Antsalo, P., Mäkivirta, A. and Välimäki, V. (2004). “Perception of Temporal Decay of Low-Frequency Room Modes”, *Audio Eng. Soc. 116th Convention*, Preprint 6083. ⏎

Kleiner, M. and Tichy, J. (2018). *Acoustics of Small Rooms*, CRC Press, Boca Raton, FL ⏎

Louden, M. M. (1971). “Dimension-Ratios of Rectangular Rooms with Good Distribution of Eigentones”, *Acoustica*, vol. 24, pp. 101–104. ⏎

Martens, W. L., Braasch, J. and Woszczyk, W. (2004). “Identification and Discrimination of Listener Percepts Associated with Multiple Low-frequency Signals in Multichannel Sound Reproduction”, *117th Convention, Audio Eng. Soc*., Preprint 6229. ⏎

Meisner, M. (2017). “A Novel Method for Determining Optimum Dimension Ratios for Small Rectangular Rooms”, *Polish Academy of Sciences, Archive of Acoustics*. ⏎

Non-cuboid Iterative Room Optimizer (NIRO), Redi Acoustics, Copyright © 2024 REDI Acoustics LLC™

Olive, S. E. (2004a). “A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 1—Listening Test Results”, *116th Convention, Audio Eng. Soc*., Preprint 6113. ⏎

Olive, S. E. (2004b). “A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 2—Development of the Model”, *117th Convention, Audio Eng. Soc*., Preprint 6190. ⏎

Olive, S. E., Schuck, P. L., Ryan, J. G., Sally, S. L. and Bonneville, M. E. (1997). “The Detection Thresholds of Resonances at Low Frequencies”,

J. Audio Eng. Soc., vol. 45, pp. 116–128.
Rindel, Jens H. (2021). "Preferred Dimension Ratios of Small Rectangular Rooms", *JASA Express Letters*, vol. 1, p. 021601.
Santillán, A. (2001). "Spatially Extended Sound Equalization in Rectangular Rooms", *Jour. Acoust. Soc. Am*., vol. 110, pp. 1989–1997.
Toole, F. E. (1982). "Listening Tests—Turning Opinion into Fact", *J. Audio Eng. Soc*., vol. 30, pp. 431–445.
Toole, F. E. (1990). "Loudspeakers and Rooms for Stereophonic Sound Reproduction", *AES 8th International Conference "The Sound of Audio"*, Paper 8-011.
Toole, F. E. (2018). *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms*, 3rd edition, Routledge Press.
Toole, F. E. and Olive, S. E. (1988). "The Modification of Timbre by Resonances: Perception and Measurement", *J. Audio Eng. Soc*., vol. 36, pp. 122–142.
Walker, R. (1992). "Low-Frequency Room Responses: Part 2—Calculation Methods and Experimental Results", *BBC Research Department Report BBC RD 1992/9*, Appendix.
Welti, T. S. (2002a). "How Many Subwoofers are Enough", *112th Convention, Audio Eng. Soc*., Preprint 5602.
Welti, T. S. (2002b). "Subwoofers: Optimum Number and Locations". www.harman.com/audio-innovations
Welti, T. S. (2004). "Subjective Comparison of Single Channel versus Two Channel Subwoofer Reproduction", *117th Convention, Audio Eng. Soc*., Preprint 6322.
Welti, T. S. (2012). "Optimal Configurations for Subwoofers in Rooms Considering Seat-to-Seat variation and Low-Frequency Efficiency", *Audio Eng. Soc. 133rd Convention*, Preprint 8748. Go to www.harman.com/audio-innovations to download color versions of the plots in this paper, Also at OR Chapter 14.
Welti, T. S. and Devantier, A. (2003). "In-Room Low Frequency Optimization", *115th Convention, Audio Eng. Soc*., Preprint 5942.

Welti, T. S. and Devantier, A. (2006). “Low-frequency Optimization Using Multiple Subwoofers”, *J. Audio Eng. Soc*., vol. 54, pp. 347–364. ⏎

CHAPTE
R 15

Headphones: The Perception and Measurement of Their Sound Quality

Sean Olive

DOI: 10.4324/9781003477495-16

15.1 THE PSYCHOACOUSTICS OF HEADPHONES

Reproduction of stereo over headphones is a fundamentally different listening experience than what is heard through loudspeakers. The topic was briefly discussed in Section 5.8.3, but it is worth repeating in more detail here. Stereo recordings are intended to be reproduced through loudspeakers in rooms—not headphones. The important psychoacoustic cues delivered to listeners through loudspeakers in rooms are fundamentally different from those presented by headphones. It should be no surprise that the timbral and spatial impressions are also quite different. This section explores why and how the impressions are different.

15.1.1 Spatial Impressions

Loudspeakers in rooms. Listeners hear a combination of the direct and reflected sounds that arrive from different directions. Together they produce impressions of direction, distance, apparent source width (ASW) and spaciousness that vary depending on the strength, direction and arrival times of the reflections. The room reflections produce mostly positive impressions of spaciousness as long the reflections produced by the loudspeaker are spectrally neutral and arrive at lateral directions where they produce the greatest interaural difference at the ears (see Section 6.1.3).

The distance of the sound stage is perceived somewhere near or beyond the locations of the loudspeakers and remains externalized and relatively stationary when the listener moves their head. Localization of the individual instruments is determined by the interchannel amplitude and time differences encoded in the recording and delivered in the direct sound from the loudspeakers. The precedence effect suppresses the influence of room reflections for localization, but not for loudness and spatial effects (Section 6.1). This phenomenon can exist in headphones only for head-tracked binaural (dummy head) recordings, or binaurally synthesized versions.

Stereo phantom images involve interaural crosstalk between the left and right loudspeakers resulting in acoustical interference (comb filtering) at the ears (see Section 4.3.1). The effect is greatest when the left and right signals are identical producing a phantom center image that has a destructive interference notch centered at 2 kHz (see Section 4.3.2).

Headphones. For headphones the listening room is removed from the reproduction, as well as any spatial or timbral benefits it added to the experience. The interchannel amplitude and time differences that established the externalized phantom stereo soundstage between the loudspeakers now become inter-aural amplitude and time differences and the soundstage is somewhere near or inside the head. There is no precedence effect information in stereo recordings, only in binaural recordings. What spatial impressions are left are completely determined by the recording and the headphone. Any sense of distance or normal soundstage will require a creative imagination and strong will to overcome the fact that the psychoacoustic cues simply do not exist. Without the proper

distance and head position cues, the featured artist front and center is often inside the head. This can either be pleasurable or uncomfortable depending on how you feel about the artist and respecting social distances. In contrast to loudspeaker listening, the sound stage doesn't remain stationary when the listener moves their head but shifts in the same direction. Often the soundstage can appear elevated or behind the listener instead of in front because the proper localization cues and head movements used to resolve ambiguous sounds in space are not available. This can be fixed in head-tracked binaural audio where the localization cues are updated as the head position changes. The result is a stationary sound stage that appears more natural and externalized (Begault et al., 2001).

15.1.2 Timbral Impressions

Loudspeakers in rooms. The perceived timbre for loudspeaker reproduction is attributed to a combination of the recording and the quality and relative levels of the direct and reflected sounds produced by the loudspeaker. The direct sound plays a dominant role in localization but also timbre. Typical forward-firing loudspeakers are omnidirectional at bass frequencies, becoming progressively more directional with increasing frequency. Highly rated loudspeakers radiate flat direct sound, so the varying directivity means that measured steady-state room curves, which are dominated by off-axis sound, will tilt gently downward. Below the transition frequency small-room resonances dictate sound quality. Figure 2.6, Section 5.1 and Chapter 14 elaborate. This is visible in the steady-state response of the loudspeaker in the room, which, above the transition frequency, can be estimated by the anechoic predicted in-room response, PIR, which contains a combination of direct and reflected sound (see Section 12.3.3).

Headphones. In headphones, the listener essentially hears only the direct sound with no contribution of the low-frequency information added by the loudspeaker's interaction with room. The perceived timbre may sound thin and bright compared to loudspeaker reproduction. This will certainly be the case if the headphone is calibrated to the current industry standard: the diffuse-field (DF) target curve. It specifies the bass be flat, which for a well-

designed loudspeaker only occurs in an anechoic chamber. We don't make recordings or listen to them in anechoic chambers, an indication that perhaps the standard is ill-conceived. This is discussed in detail in Sections 15.5.2 and 15.6.

Phantom center. The phantom center experience in loudspeakers does not exist in headphones since there is no interaural crosstalk between the left and right ears. The steep notch that can occur at 2 kHz while sitting in the sweet spot between two loudspeakers is not present in headphone listening and should produce a more neutral timbre (see Section 4.3.2). The Dolby Atmos mixing process allows for optional binaural processing of LCR channels for headphone playback. It is intended to simulate the locations and timbre of loudspeakers over headphones including the interaural crosstalk associated with center phantom images and the 2 kHz notch. Many mixers prefer to leave it turned off (King et al., 2023).

Regardless of any timbre improvement, the headphone experience is different from the one heard through loudspeakers.

15.1.3 Reduced Sensitivity to Resonances

Toole and Olive (1988) found that headphone and anechoic listening yielded the highest detection thresholds for resonances, meaning that listeners are relatively forgiving of headphone resonances. Fortunately, in-ear headphones are generally free of resonances over most of the frequency range. On- and around-ear designs have more complications. Most designs are susceptible to variations in bass level due to air leaks. This can have significant consequence on the experience, since the quality and quantity of bass accounts for about 30% of the listener's preference (see Sections 11.1 and 11.2).

15.1.4 Lack of Tactile Experience in Headphones

Several studies have reported the impression of that headphones are perceived as less loud than loudspeakers when identical signals are presented at the ears. Rudmose (1982) wrote a paper titled "The Case of the Missing 6 dB" which has since been largely attributed to methodological

errors. Völk and Fastl (2011) showed that if the signals produced from the loudspeakers are binaurally presented through calibrated headphones using individualized HRTFs, the loudness impressions between the two will be the same. However, headphone reproductions will be missing any mechanical whole-body and airborne vibrations that exist from the loudspeaker reproduction.

Simon et al. (2009) compared automotive sound systems subjective evaluations using both in situ and binaural reproductions over headphones. Listeners reported there was less bass in the binaural reproduction even though the signals at the ears were identical in both methods. When the tactile or whole-body vibrations produced by the audio system were measured and then simulated by a shaker in the seat, the bass in the headphones was judged to be equivalent to the in situ experience. The tactile effect produced an equivalent effect of adding 3 dB of bass to the headphone. Simulating the right amount of vibration to achieve equivalency with the loudspeaker experience depended on bass in the music and the mass or weight of the listener. There is a limit to how much whole-body vibration a listener will tolerate, and beyond that listeners prefer it to be turned off.

15.1.5 Immersive Audio Over Loudspeakers vs. Binaural Headphones

The previous sections focus on the psychoacoustics of stereo reproduction through loudspeakers. Stereo is a spatially limited experience through loudspeakers that is even more compromised with headphones. Multichannel loudspeaker reproduction seriously upgrades the spatial experience as discussed in Chapter 8. *What happens when the headphone stereo experience is upgraded to a binaurally rendered version reproduced through headphones?*

The author has at home a high quality 7.2.4 loudspeaker system and well-designed headphones with the ability to compare immersive film and music such as Dolby Atmos or Apple Spatial Audio through both systems. When switching from one to the other, the headphone experience is often

underwhelming and disappointing. While the spatial impressions over headphones are an improvement over stereo, the sense of direction, envelopment, and space is more precise and realistic through loudspeakers. In the case of the loudspeaker reproduction, I am hearing the sounds through my own personal set of HRTFs—not the generalized HRTFs used by the binaural renderer. The binaural version is further compromised since the headphones have not been calibrated to my ears and dynamic head tracking is missing. The timbral impressions are also different with the loudspeaker reproduction having more tactile impact that shakes the walls, pant legs and body—something that is absent in headphones. The promise is that personalized HRTFs, calibrated headphones, head-tracking and some tastefully added vibration (through an added subwoofer or shaker) will bring the headphone experience closer to what is felt through loudspeakers. Until then, the headphone immersive experience is a compromised experience compared to the loudspeaker one.

15.1.6 Single Stimulus Comparisons, Auditory Memory and Adaptation

Much has already been written about fallacy of single stimulus evaluations where a product reviewer takes the product home and over weeks comes up with an opinion (Section 1.4). Adaptation, reduced sensitivity due to limited auditory memory and the lack of available anchors and references together with sighted/expectation biases can produce an opinion that deviates far from the norm—an outlier. At best the opinion may be accurate but how can you know or trust it is accurate? The virtual headphone method (Section 15.6.2) eliminates these nuisance variables facilitating sensitive and repeatable judgments of headphone sound quality. While sensitivity and accuracy in measurements are necessary to get to the truth, in practice the audible differences and their influence on preference reported in our research may be beyond what the average listener or reviewer can discern using less sensitive methods.

15.1.7 Summary

Headphones produce a fundamentally different stereo experience from loudspeakers in both spatial impression (distance, direction and spaciousness) as well as perceiving low and medium Q resonances where reflections from the room enhance audibility. Fortunately, most modern headphones, particularly in-ear models are well-behaved in controlling resonances. Resonances tend to be problematic at higher frequencies, particularly for AE/OE models where the diaphragms begin to break up and there are complex acoustical interactions with the pinnae of the listener.

What remains are mostly spectral balance and broadband spectral trends in frequency response that can influence the perception of timbre. The coupling of the headphone to the ear can affect this variable, particularly at low frequencies where leaks occur. And since the fit and coupling of the headphone to the ears depends on the varying shape and size of the individual's head and ears, opinions about headphone sound quality vary widely.

The challenge in good headphone design as well as perceptual research related to their sound quality is to overcome these variables. Only when the same signals are delivered to listeners ears can we expect them to agree on what sounds good. And only when the objective measurements capture those signals can be we begin to correlate and predict their subjective impressions.

15.2 HEADPHONES: THE PRIMARY DEVICES THROUGH WHICH WE LISTEN TO REPRODUCED SOUND

The rapid growth of headphones witnessed since early 2000 can be traced to technological disruptions: introduction of the MP3 player like the iPod followed by the smartphone, and the growth of digital media streaming services. Music and entertainment became mobile and could be enjoyed everywhere. Headphones became the primary device through which entertainment was experienced (Statistica, 2023) and sales increased as their form factor became smaller and wireless. Even in the home, up to 60%

of audio is heard through headphones vs. loudspeakers (Bull and Duvall, 2023).

Headphones provide many benefits not normally associated with loudspeakers: they are small, portable, cost-effective, power-saving devices that provide privacy while not disturbing others. The best noise cancelling models can provide up to 30 dB of isolation from disturbing background noise (Rtings, 2024). Some models can compensate for moderate hearing loss, provide hearing protection from loud noises, and increase speech intelligibility in noisy environments by reducing the background noise while increasing the gain of the person talking to you (Apple, 2024).

Headphones can offer excellent fidelity and performance for much less cost compared to loudspeakers. For as little as $20 a pair of (well chosen) wired in-ear (IE) headphones can produce a smooth, flat, accurate frequency response across the audio bandwidth up to 110 dB with lower distortion than many loudspeakers costing thousands of dollars.

Despite these advantages, some people refuse to wear headphones for reasons related to comfort, isolation from the outside world, the lack of tactile feedback from bass, and in-head localization: the unnatural feeling that the stereo image appears to be near or inside your head.

Loudspeakers provide a better sense of distance, depth and externalized sound. While many audiophiles wax poetic about the stereo imaging of their headphones, Silver (2024a) argues the psychoacoustic cues required for a good spatial audio experience over headphones (e.g. dynamic HRTFs) are simply not available. On the other hand, headphones can produce an enhanced sense of intimacy with the musicians or narrators that loudspeakers cannot produce. *Is a heightened sense of intimacy worth the other spatial trade-offs associated with headphones?*

Kallinen and Ravaja (2007) compared subjective and psychophysical responses of listeners listening to news on headphones vs. loudspeakers. They found headphone listening elicited more positive emotional responses and higher attention, and listeners preferred to listen to news on headphones over loudspeakers. In another study, Lieberman et al. (2022) found in-head localization promoted feelings of connectivity, empathy and engagement

with the narrator and more likely agreed with them. They recommend that advertisers and podcasters should target platforms where the primary audience are headphone users. Headphones are apparently effective for brainwashing.

Headphone reproduction of binaural or spatial audio is a different, more elevated experience when the content is rendered with personalized head-related transfer functions (HRTFs). HRTFs define interaural time and level differences and spectral cues that allow us to localize a sound source in space.

Musicians prefer playing over headphones when the monitoring feed is binaural vs. stereo, and head-tracking is used (Matteo and Truchet, 2023). All the missing psychoacoustic cues that Silver discusses are available and many of the spatial artifacts associated with stereo reproduction disappear. The sense of social isolation with headphones is solved with Auracast (Scarrott, 2023). It allows music streams to be broadcast via Bluetooth to an unlimited number of headphones so that everyone at the party can share the same experiences. Transparency mode, a feature that uses the microphone in the headphone to feed in outside sounds allows the user to have conversations with others without removing their headphones. It also allows the user to hear their own voice with a reduced occlusion effect.

15.3 TYPES OF HEADPHONES

Headphones come in different form factors but can be broadly categorized as having a headband or not. Figure 15.1(a)–(c) are examples of headband headphones, and (d)–(f) have no headband. Earphones are any headphone that is inserted into the ear canal with a foam or silicon tip that can be sized to provide a comfortable snug fit. The most common type is the in-ear (IE) headphone (sometimes referred to in professional audio circles as in-ear-monitors or IEMs) shown in (d). Earbuds (e) sit in the concha and rely on a snug fit to secure themselves in place. Often, they slip out when bending over falling into a gutter or crushed by a bus when running across the street. Manufacturers of earbuds have repeat customers for life. People who have

active lifestyles and don't like their ear canals occluded might choose an open ear type such as (f). It is secured with a hook behind the ear.



FIGURE 15.1 *The top row shows examples of headband headphones: (a) an around-ear (AE) open back professional headphone, (b) an on-ear (OE) closed back headphone and (c) a closed-back wireless AE headphone. The bottom row shows three examples of wireless headphones without headbands: (d) is an in-ear (IE) type with tips that insert into the ear canal, (e) is an earbud type that sits in the concha and (f) is an open ear design held in place with a hook that goes around the back of the ear.*

Headphones are either passive (requiring an amplifier) or active with the amplifier, equalization and signal processing built in. The downside is they require a battery that can last 20–30 hours before requiring recharging.

Passive headphones require a cable to connect and tether to an amplifier, whereas wireless models use Bluetooth or Wi-Fi to send the signal from the source.

IE headphones are the most popular type accounting for more than 49% of unit sales, followed by AE (30%) and OE (14%), with earbuds (7%) being the least popular type sold (Bems, 2023a). Wireless headphones with active noise cancellation (ANC) have become a must-have feature for travelers, commuters, office workers or anyone who requires privacy and isolation from noisy environments.

The most popular transducers found in headphones are dynamic drivers due to their relative low cost to performance ratio. Magnetic planar drivers are found in more upscale audiophile headphones with some added size and weight. Electrostatic transducers are found in the most expensive models and require high-voltage amplifiers. You are not likely to find audiophiles wearing these in business class flights. IE headphones also use dynamic drivers that are sometimes combined with balanced armatures (BA) that are used to reproduce higher frequencies due to their limited bass output. More recently, solid-state microelectromechanical systems (MEMS) transducers such as XMEMS (www.xmems.com) offer many advantages because of their small size, high linearity and accurate phase up to high frequencies, and consistent quality in manufacturing (Wykes, 2024a).

15.3.1 Headphones With Headbands

Around ear (AE) or circumaural headphones have cups with transducers that enclose the ears and come in two versions: closed back or open back. Closed-back models typically rely on a good seal around the ear to produce adequate bass, which also provides some isolation from background noise, and keeps sound inside from disturbing others. When the seal is poor, an air leak occurs, often creating a significant roll-off in the bass beginning as high as 500 Hz. It can also have adverse effects on active and passive noise performance. Therefore, a good fit and seal is a critical factor when choosing a closed-back headphone and something you should consider before purchase. The best closed-back designs are designed to

accommodate a range of sizes and shapes of ears and heads. In some cases, a leak is incorporated into the design, relaxing the fit requirement.

In open-back designs the space behind the transducer is fully exposed or in semi-open designs partially exposed to the outside. Sound can freely pass inside to the outside and vice versa. This makes open-back designs more suited for solitary listening in quiet spaces. Open-back designs can produce low bass but typically less bass than closed-back designs. Because the fit and seal have little impact on the bass performance, the frequency response tends to very consistent across listening sessions and among different listeners. This makes them excellent choices for psychoacoustic experiments. Many audiophiles seem to prefer open-back designs for their alleged impression of a more open and spacious sound compared to closed backs. This could possibly be related to hearing the acoustics of the room they are listening in. Whether there is any psychoacoustic evidence for these claims is a matter of debate.

On-ear (OE) or supra-aural headphones have smaller cups and cushions that sit on the ear rather than around it. This smaller form means the headphones are typically lighter, smaller and more portable. The trade-offs are that they are often less comfortable over periods of time because the clamping force from the headband puts all the pressure on the ear itself rather than distributed over a wider area around the ear as in the AE type. The lack of a tight seal around the ear means there is poor isolation from background noise and low bass can be a challenge. Comfort, performance and noise isolation may explain why these headphone types are the least popular, and typically aimed at lower cost products for kids.

15.3.2 Headphones Without Headbands

Headphones that don’t require a headband are much smaller and rely on an ear tip inserted into the ear canal to hold them in place or they are shaped to fit inside the concha, often secured by a flange that presses against the inside of the helix, or a hook placed over the top of the ear. Comfort, bass performance, and isolation from background noise are factors that will likely determine which type best suits your needs.

In-ear headphones (with ear tips) are designed to be inserted into the ear canals and come with a range of different sized silicon or foam tips to accommodate different diameters of ear canal openings. They provide the most bass of any type provided you have a good seal. They come in both passive wired and active wireless versions and two different form factors. The most widely sold and known example are the Apple AirPods. Transparency mode can ease the impression of isolation.

Earbuds (no tips). Many people prefer not to have their ear canals obstructed and find in-ear headphones uncomfortable and prefer not to be isolated from the outside world. Earbuds sit in the concha of the ear avoiding occlusion of the ear. However, as mentioned above the security of the fit can be problematic especially if the user is active. By design these headphones have less passive noise isolation and less bass than IE headphones. However, earbud models are now available with ANC to provide more attenuation of background noise.

Open-ear headphones are designed for maximum openness and transparency with the intent that the user can listen to sound while being fully aware of their surroundings. They are good choices if you are active outdoors and want to hear oncoming traffic or any situation where you need to be aware of sounds around you. The main issue with these headphones is they generally lack bass, and you tend to turn them up in the presence of competing background noise. This raises a question whether certain headphone types are more prone to producing hearing loss than others. For people who want better fit, deeper bass, the ability to switch from isolation to full transparency an IE headphone with transparency mode is a good option.

Bone conduction headphones use vibrations from a transducer that rests on the cheekbone or behind the ear that are sent through the skull directly to the cochlea, bypassing the ear drum. For people with hearing deficiencies this can be of great benefit, but for normal hearing listeners the fidelity is noticeably degraded compared to traditional air-conduction headphones. Bone conduction headphones offer no passive or active isolation from background noise, but they are relatively immune to pickup of background

noise. For this reason, bone conduction microphones can improve the intelligibility of voice in high background noise situations when supplemented with the regular microphone feed.

15.3.3 Headphone Type and Hearing Loss

Headphones connected to portable device can produce SPLs as high as 120 dB that can cause temporary or permanent hearing damage with increased exposure times (Hammershøi et al., 2008). Hearing damage from headphones is particularly a concern among teens and young adults who wear headphones for long periods of time and are often unaware of the potential risks (Muchnik et al., 2012). The rate of hearing loss in teens is estimated to be 30% higher than it was before 1990 when headphones were less popular.

Fasanya and Strong (2019) surveyed 280 college students who frequently wear headphones. Forty-four percent of daily users did not believe headphones cause hearing damage despite reporting symptoms normally associated with hearing damage: 25% experienced ringing in their ears, 33% experienced trouble sleeping and 23% believed that headphones had negatively affected their academic performance.

Dobrucki et al. (2013) measured hearing loss of 81 young people (16 to 25 years of age) to study the influence of headphone type on hearing loss. The subjects claimed to listen frequently to loud music on headphones and were classified into four categories: young classical musicians, front-of-house/public address (FOH/PA) sound engineers, sound engineers working in recording studios, and general users of headphones not involved in the entertainment industry. Not surprisingly, the FOH/PA engineers had the highest average hearing loss regardless of headphone type, with 9 dB more hearing loss at 4 kHz than the next closest group. The average hearing loss at 4 kHz was observed for in-ear headphones (9.2 dB), with less hearing loss for open (3.9 dB), closed (2.9 dB) and semi-open (0.1 dB) designs. Hodgetts et al. (2009) found increased background noise and activities like exercise tend to lead to higher preferred headphone levels. This suggests that headphone with good passive and active attenuation of background

noise can lead to lower preferred playback levels and prevent noise-induced hearing loss.

Headphone companies like JBL (2024), Puro Sound Labs (2024) and Apple (2024) design headphones aimed at children and teens and limit the SPL to 85 dB to prevent hearing damage. JBL provides an app that allows parents to monitor and limit the SPL as well as the exposure times. The Apple AirPods Pro 2 offers active hearing protection in transparency mode where the SPL during a concert or other loud event can be reduced by as much as 18 dB with better fidelity than a passive hearing protector. With transparency mode off background noise can be reduced up to 22–30 dB, which can lead to lower and safer preferred listening levels. These are examples where headphones prevent hearing loss instead of causing it.

15.4 HEADPHONE TEST FIXTURES

Headphones are closely coupled to our head and ears, and their positioning, fit and acoustic seal around the ear can significantly influence their sound. For this reason, headphones are designed and tested on headphone test fixtures that aim to simulate the anatomical features of an average human head and ear usually including the pinna, ear canal and ear drum.

Figure 15.2 shows four different headphone test fixtures commonly used for measuring headphones. For measuring IE headphones only, the ear canal is required and a test fixture like the GRAS 43AG (Figure 15.2(a)) or the Brüel & Kjær (B&K) Type 4620 (Figure 15.2(d)) are needed. Both fixtures include an ear canal simulator intended to acoustically simulate the volume, length and input impedance of an average human ear canal. An anthropometric pinna is an option for the GRAS but integrated into design of the B&K 4620. Different-sized pinnae are available with different hardness or compliance measured in Shore units. The hardness or stiffness of the pinna as well as its shape or size can affect how well a headphone will seal around the ear (Welti, 2015). A microphone is placed where the ear drum terminates the canal and simulates its acoustic impedance.

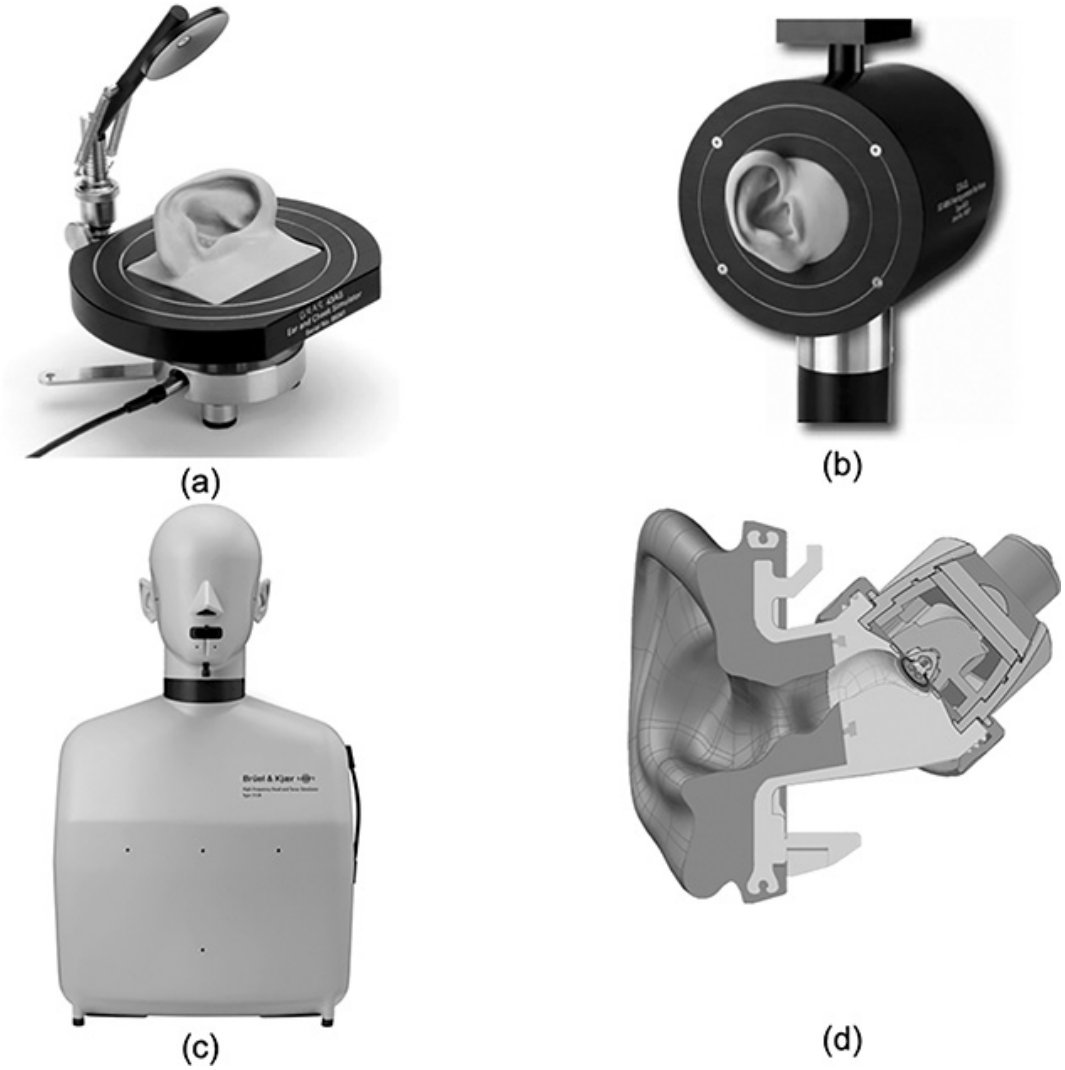


FIGURE 15.2 *Two examples of headphone test fixtures based on the ITU-T Type 3.3 ear simulator are (a) the GRAS 43AG and (b) the GRAS 45CA. Fixtures based on the ITU-T Type 4.3 include (c) the B&K Type 5128 head and torso simulator (HATS) and (d) the B&K Type 4620 ear simulator, which is included in (c) but can be purchased separately.*

Headphones with headbands require a test fixture that can support them while under test. How well the headphone fits and is coupled to the artificial head and ear affect its acoustical performance. The better the dimensions, shape and materials of the head, cheek, pinna and ear canal match those of an average human, the better the headphone's sound will

correlate to what is measured and heard on humans. Figure 15.2(b) and (c) are examples of test fixtures that can be used for testing headphones with headbands, and there are several other solutions available from GRAS, B&K and Head Acoustics that conform to Type 3.3 and Type 4.3 international standards These are briefly discussed below.

15.4.1 Type 3.3 Ear Simulator

The Type 3.3 ear simulator is described in IEC: 60318–4 (2010) standard. It is often referred to as a "711" coupler named after the original IEC 60711 standard (1981) from which it evolved. The ear simulator is intended to approximate the volume and acoustical properties of an average human ear canal over a limited bandwidth of 100 Hz to 10 kHz. The accuracy of measurements beyond those frequencies is problematic and prone to misinterpretation.

15.4.2 Type 4.3 Ear Simulator

In 2018 B&K released the first Type 4.3 ear simulator known as Type 4620. It provides a more accurate acoustic representation of the human ear canal from 20 Hz to 20 kHz while maintaining compliance with the Type 3.3 standard. A large study was conducted where contrast-enhanced MRIs were taken from 44 adult Caucasian subjects (12 females and 32 males) with normal hearing ranging in age from 24 to 55 years (Darkner et al., 2018). The average geometry and acoustic impedance of the canals were determined, and this was the basis for the Type 4.3 ear simulator specified in ITU-T P. 57 (2021).

15.4.3 Head and Torso Simulators

Head and torso simulators (HATS) are acoustic manikins that include an artificial head, torso, and the external ear (pinna, ear canal and drum). HATS specifications are defined in the ITU-T Rec. P. 58 (2023), IEC 60318–7(2022) and ANSI/ASA S3.36 (2018) standards.

HATS are used for measuring and recording sound fields of interest that include the acoustic interaction effects with a human-like manikin. The

torso produces a reflection off the shoulder (Blauert, 1983), that creates a 5 dB notch around 700 Hz, and a 20 dB shadowing effect at higher frequencies when the sound source is positioned below the torso, for example, door speakers in cars. An optional mouth simulator includes a calibrated loudspeaker for simulating human speech used to measure the performance of microphones in headsets and phones. More portable tabletop versions of HATS minus the torso are available for headphone-only applications. Tabletop and full versions of HATS are available from GRAS, B&K and Head Acoustics.

15.4.4 Headphone Measurement Challenges

The ideal headphone measurement is accurate, reproducible and agrees with measurements made on an average human: the closer they match, the higher the correlations will be between the objective and subjective measurements. However, there are many errors that prevent a perfect match including the accuracy of the test fixture, how the headphone couples to it, and differences in how the headphone couples to a human. These errors are discussed below.

One source of error is the accuracy of the test fixture itself. Accuracy in this sense is how well its acoustical transfer impedance approximates that of a human ear. The transfer impedance will affect how it interacts with the acoustic load of the headphone and ultimately the frequency response measured at the microphone. The acoustical transfer characteristics of Type 3.3 and 4.3 ear simulators are defined in the standards and their effect on the measured frequency response on different IE headphones reported by Mansilla (2020) and Olive et al. (2024) in Section 15.6.9. Measurement differences between Types 3.3 and 4.3 for AE and OE headphones are discussed in the following section.

The reproducibility of headphones measurements on both test fixtures and humans has been well documented (Møller et al., 1995b; Kulkarni and Colburn, 2000; McAnally and Martin, 2002; Völk, 2014; Schärer and Lindau, 2009; and Welti, 2015). The positioning and coupling of the headphone on the ears can produce significant variations in the measured

response, particularly at frequencies below 500 Hz and above 3 kHz. Closed-back headphone designs typically require the cushions to properly seal around the ear to reproduce good bass, which can be a challenge for many test fixtures and humans depending on the size and shape of their heads/ears and how well the cushions fit and seal around the ear.

The anatomical variability in human heads and ears means that there can be significant variability in measurements of headphones made on humans. The stiffness of the pinna and how easily it can be formed or flattened by the headphone cushion can affect how well it seals. When the cups are large and easily fit around the pinna the shape and contour of the cheek and jaw can affect the seal and cause a leak. Hair and glasses can also introduce leaks. What the listener hears (a lack of bass) may not be reflected in the test fixture measurement, and the correlations between the two become weaker. The error variance associated with headphone positioning is usually dealt with by taking the average of several measurements with repositioning in between each one.

For IE headphones carefully selecting the right size and shape of ear tips can achieve a reliable seal and repeatable measurements are easier to achieve. The main challenge is maintaining a constant insertion depth as this effectively changes the ear canal's length, volume and transfer impedance (Johansen et al., 2024). The change is visible as a frequency shift in the resonance related to the length of the canal: deeper insertions will shift the resonance a few kHz higher around 8 kHz.

Section 15.4.6 discusses differences in measurements of headphones made on test fixtures vs. humans and the implications the errors have on interpreting how a headphone sounds based on a measurement.

15.4.5 Effect of Test Fixture on Headphone Measurements

Headphone measurements are widely found on both professional and enthusiast reviewers' websites. The measurements are made on both industry standard test fixtures and imitation versions that are best ignored.

How similar are headphone measurements made on different test fixtures that are all compliant to industry standards? One might expect them to

produce similar results, but how similar are they in practice?

To explore this question, our research team measured 20 different models of headband-type headphones on five different test fixtures. The headphones included 17 AE and 3 OE models with 17 being closed-back and 3 open designs. Each model was measured five times with repositioning it on the fixture between measurements. An average frequency response was calculated for each test fixture based on 200 measurements (20 headphones × 2 channels × 5 repeats).

The five test fixtures were:

1. GRAS 45CA/MOD
2. GRAS 45CA/KB5001/02 Anthropomorphic pinna
3. Audio Precision AECM206
4. KEMAR 45BC/KB5001/02 Anthropomorphic pinna
5. B&K Type 5128

Fixtures 1 and 2 were identical except for the pinna. Fixture 1 is a Type 3.3 ear simulator with a modified (MOD) version of the GRAS KB0070/71 pinna that meets the ITU-T P. 57 (2021) Type 3.3 standard. The modification produces headphone leakage effects that better approximate what is measured on humans (Welti, 2015). (See OR Chapter 15.) Fixture 1 was used for the Harman research between 2013 and 2019, so there was an interest in comparing it to headphones measured on other fixtures.

Shortly after Welti published his research, GRAS released the KB5001/2 anthropomorphic pinna used on Fixture 2. The profile and softness produced leakage like that measured on humans, and it improved measurement reproducibility.

Fixture 3 was an Audio Precision AECM206, also a Type 3.3 ear simulator. Its form factor was clearly inspired by the GRAS 45 CA.

Fixture 4, GRAS KEMAR 45BC has the same pinna and ear simulator as Fixture 2 but includes a head and torso.

Fixture 5 is a B&K Type 5128 and is the only Type 4.3 ear simulator in the group. The other important distinction here is that Fixtures 4 and 5 are both HATS, and Fixtures 1, 2 and 3 are flat plates with pinnae attached.

Figure 15.3(a) shows the average frequency response of the 20 different headphones measured on each of the five test fixtures. The measurements have been normalized at 500 Hz. Figure 15.3(b) shows the same measurements relative to Fixture 1, our reference fixtures used throughout much of the Harman research. In this we show the difference between Fixture 1 and the other four fixtures.

Fixture 2 (GRAS KB5000) has the closest match to Fixture 1 (GRAS 45 CA MOD)—our reference test fixture—other than 1–2 dB attenuation effects below 200 Hz. Since the ear simulators are otherwise identical this is not surprising. The differences can be attributed to more leakage produced by the KB5000 pinna.

Fixture 3 (AP AECM206) produces measurements that closely match those of Fixture 2, not surprising given this was clearly its design goal. There are some departures between 3 kHz and 5 kHz where Fixture 3 has 2–3 dB less level relative to Fixture 2.

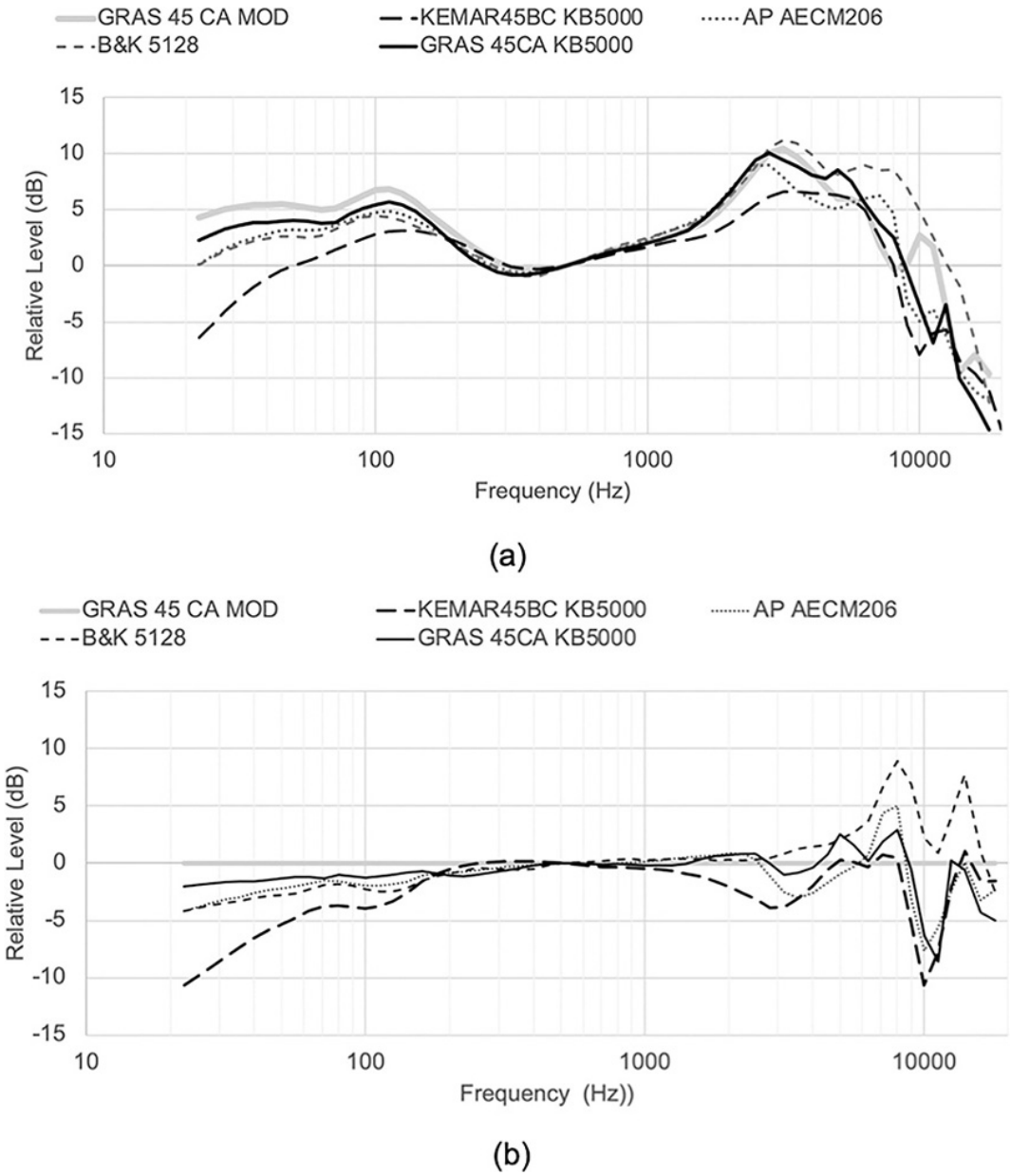


FIGURE 15.3 *(a) The average frequency response of 20 different AE and OE headphones measured on five different test fixtures. (b) The differences between the above fixtures relative to the Fixture 1 (GRAS 45 CA MOD) used for Harman research. MOD is a modification to the GRAS pinna to better match human leakage effects in headphones (Welti, 2015).* ⏎

Fixture 4 (KEMAR 45BC/KB5000) deviates furthest from Fixture 1 even though it shares the same coupler and pinna of Fixture 2. It has noticeably

more leakage below 200 Hz dropping 10 dB below the reference, and an attenuated high-frequency response above 2 kHz. The leakage effects shown here are common to headphones measured on a head vs. a flat plate. The high-frequency differences are not as easy to explain. One possibility is that headphone is positioned at a slight angle on the manikin compared to the flat plate, and this is accounts for the reduction in high-frequency level.

Fixture 5 (B&K Type 5128) produces measurements that significantly deviate from Fixture 1: there is a 5 dB low-frequency roll-off beginning at about 500 Hz and significantly more output above 3 kHz. The differences can be attributed to differences in the acoustical transfer impedance of the Type 4.3 coupler compared to the Type 3.3, and also leakage effects that tend to be larger on artificial heads compared to flat plates.

In summary, there are significant differences in headphone measurements made on different test fixtures even those that meet the same standard (Fixtures 2 and 4). The flat plate fixtures produce less leakage at low frequencies than measurements made on HATS. The Type 4.3 fixture produced measurements with elevated high-frequency levels compared to Type 3.3 fixtures. Headphones designed to a target curve on Fixture 1 (our reference) will measure differently on another test fixture, and therefore the same target cannot be applied. Otherwise, it could lead to a misinterpretation of the measurement and how the headphone sounds.

An important question is which test fixture produces measurements that best match those made on humans. That question is addressed in the next section.

15.4.6 Comparing Headphones Measured on Humans vs. Test Fixtures

In the previous section, headphones were measured at ear drum reference point (DRP) where the microphone is located. The equivalent measurement for a human requires a small probe microphone to be inserted down their ear canal until it just barely touches the ear drum. The procedure is time-consuming and uncomfortable for the subject with the risk of causing damage to the ear drum. These measurements are rarely performed except

in clinical research settings. Another challenge is the measurements are notoriously unreliable at high frequencies where small changes in microphone position can produce 10–20 dB differences.

Given these risks, we compared headphone measurements made at the blocked entrance of the ear canal for both the test fixtures and humans. With the ear canals removed from the measurement the observed differences between fixtures and human would be related to anatomical differences of the head and pinna—not the ear canal.

Møller et al. (1995a) measured the transfer characteristics of 14 headphones on 40 human subjects at the entrance of their blocked ears and found the frequency response was characterized by smooth structured fluctuations at low frequencies with low variation among subjects up to 3 kHz where it increased with frequency. At 8 kHz the intersubject standard deviations were 3–5 dB as the frequency of the resonances shifts among individual listeners. Comparisons between headphones measured on humans vs. test fixtures were not made as they were mostly interested in finding the best measurement approach for individual equalization of headphones for binaural reproduction.

The author used a similar approach to measure nine different AE headphones on 15 human subjects and compared the same measurements made using four of the same test fixtures discussed in Section 15.4.5. Measurements were made at the blocked entrance of the ear canal for both humans and test fixtures using a small microphone embedded in a foam plug. The headphones included five open-back and four closed-back models each measured five times with different reseats between measurements.

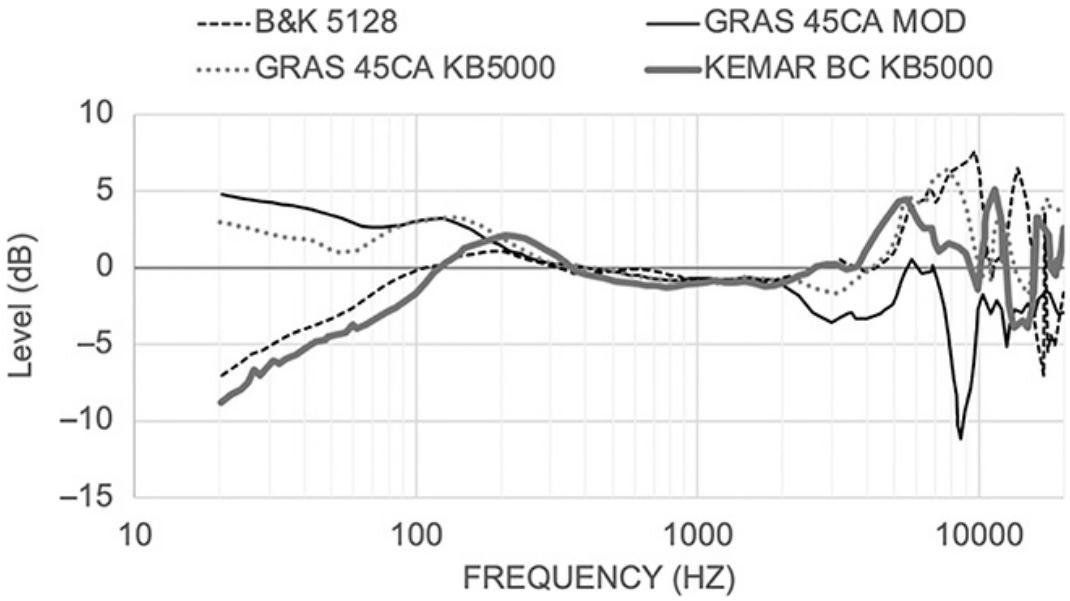


FIGURE 15.4 *The average difference between nine headphones measured on 15 humans vs. four different headphone test fixtures. The largest differences occur below 300 Hz and above 2 kHz caused by anatomical differences and how the headphone couples to the ear.* ⏎

Figure 15.4 plots the average difference between the nine headphones measured on humans vs. each of the four test fixtures. The following observations can be drawn:

- There is generally good agreement between the human and test fixture measurements between 300 Hz and 2 kHz.
- *Leakage*: The two flat plate test fixtures (GRAS 45CA with MOD and KB5000 pinna) tend to underestimate low-frequency leakage on humans by 2–4 dB and the HATS (B&K 5128 and KEMAR) tend to overestimate it by 4–5 dB. KEMAR has the largest overestimation of leakage, and GRAS 45CA MOD has the largest underestimation of headphone leakage on humans.
- *Above 2 kHz*: There are considerable differences between test fixture with ±5 dB errors or more depending on the fixture. The GRAS 45CA

MOD has the largest error overall generally underestimating the level produced by the headphone on humans between 2 and 20 kHz.

- *Overall*: The B&K Type 5128 has the closest agreement with the human measurements between 60 Hz to 4 kHz but generally overestimates the level above 4 kHz compared to the human measurements.

It is important to remember that this study does not include the ear canal. The results are not relevant to measurements of IE headphones where the pinna and head are bypassed, and the ear canal acoustics are the dominant factor. Studies of human ear canals and their objective and subjective effects on IE headphone performance are not readily available because of the methodological challenges and invasiveness of measuring signals at the human ear drum.

Jønsson et al. (2018) measured the acoustic impedance of ear canals of 32 adults both male and females from a wide range of ages. They found no statistically significant differences related to age or gender. However, they did find some significant differences when comparing the impedances to that of the Type 3.3 “711” coupler.

> At low frequencies the level of the 711 coupler was slightly higher, and in the midfrequency area it underestimates the mean human with just over one standard deviation, and at higher frequencies above the half wavelength resonance the standing wave pattern was quite different.

The data from this study was used to design the Type 4.3 B&K Type 4620 ear simulator that is included in the B&K Type 5128. For now, this is the best estimation of the acoustic impedance of an average human ear canal.

15.4.7 Reproducibility of Headphone Measurements on Humans

The ideal headphone consistently delivers the same sound to the listener regardless of the size and shape of their heads and ears, and, within reason,

how they position it on their ears. Its reproducibility can be quantified by measuring the headphone several times on different subjects and calculating the variation of the mean response.

Figure 15.5 shows the intersubject standard deviation in dB vs. frequency for 9 different headphones measured on 15 subjects. The standard deviation is calculated across all subjects and repeated measurements for both left and right channels.

The standard deviations increase below 400 Hz varying from 2.5 to 9 dB at 20 Hz. Above 1 kHz the standard deviations rise as high as 8 dB above 8 kHz. One model has unusually high standard deviations in response from 1 to 3 kHz.

As expected, the variance in response is much higher for closed back headphones compared to open backs with one exception illustrated with an arrow. This closed-back model (DCA Aeon Noire) has reproducibility below 200 Hz that is almost as good as open backs. Not shown in this data is the distribution of responses measured across the 15 listeners for closed-back models. For some listeners the headphone seals well and produces excellent bass, and for other listeners it leaks and produces little or no bass. Because the headphone cannot consistently deliver the same signals to all the listeners, the quality of listening experiences will vary, as will their opinions of how good the headphone sounds.

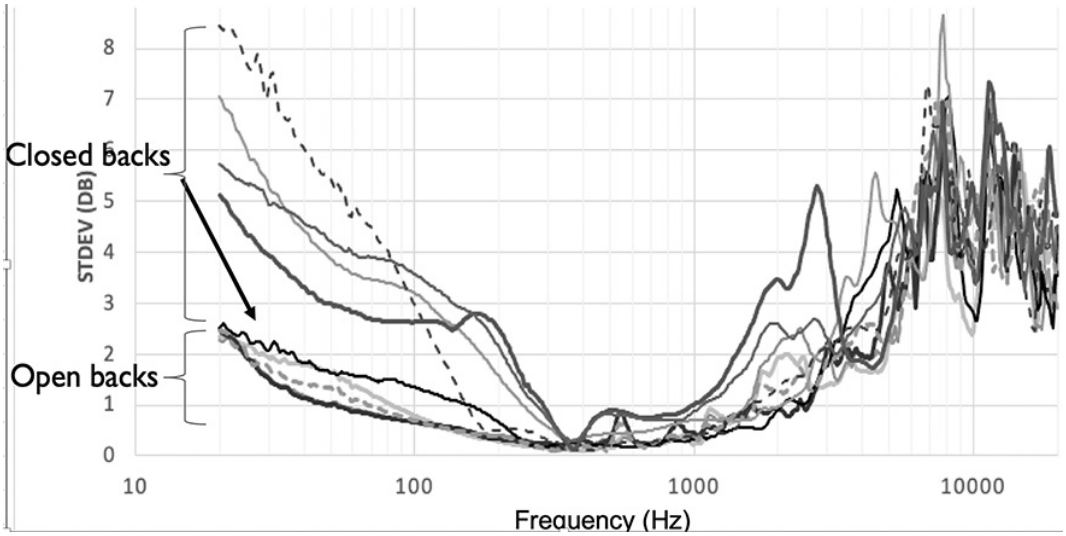


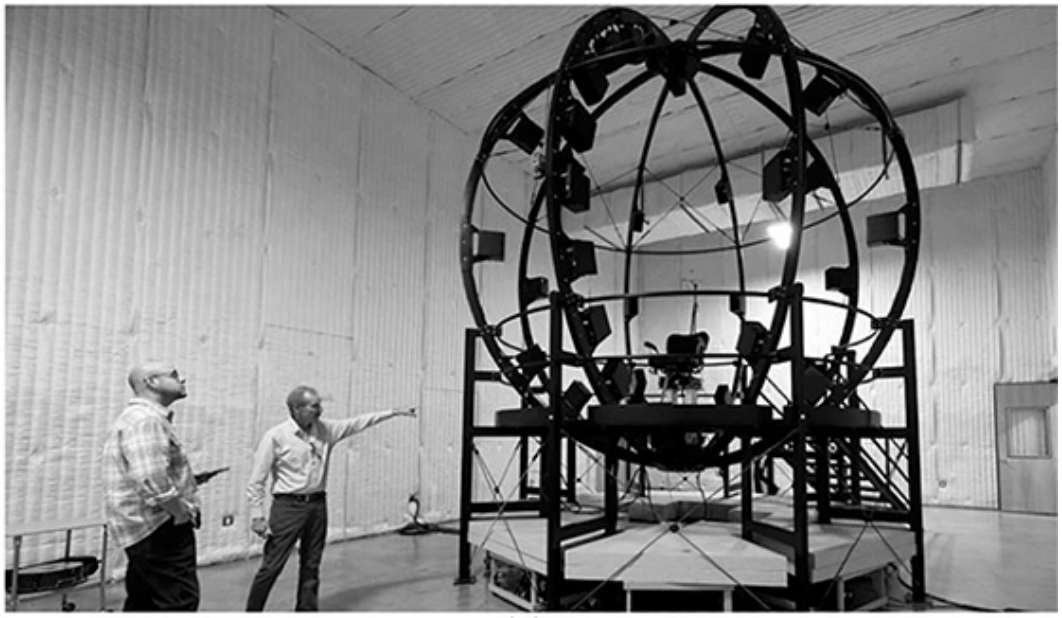
FIGURE 15.5 *The intersubject standard deviation in frequency response for nine headphones measured at the blocked ear canal entrance of 15 listeners. The closed-back models have higher variance at low frequencies due to leakage effects. The exception is the closed-back model (DCA Aeon Noire) indicated by the arrow.* ⏎

On the other hand, open-back models have a built-in leak, and with good design can deliver the most consistent frequency response across the widest range of listeners. This makes them excellent candidates for psychoacoustic experiments where the goal is to deliver the same signals to each listener so that the results are reproducible and traceable to headphone measurements made on test fixtures.

Good reproducibility of audio across the targeted customers should be an important design goal for consumer and professional headphones. But how can such a design goal be validated during the design and testing? One approach is to measure the headphones on a variety of different-sized human ears and heads as done in this study. However, finding a group of humans who are always available for testing and are anatomically representative of the population is easier said than done. A more practical approach is to build several artificial manikins whose anthropomorphic features are representative of a target human population. This was one of the goals of the Harman Hyperion Project described in the next section.

15.4.8 The Hyperion Project

In 2019, Harman built the Hyperion Sphere, a research lab designed to measure free-field HRTFs of humans and their anthropometric data via three-dimensional scans of their heads and ears (Miller, 2024). To date over 350 adults from a diverse sample based on gender, race and age have been measured, a significant addition to other libraries such as the Austrian Academy of Science Acoustics Research Institute (2024), SADIE (2004), and Listen (2002).



(a)



(b)

FIGURE 15.6 *(a) The Harman Hyperion Sphere consists of 34 calibrated loudspeakers and four subs used to measure individual HRTFS and reproduce immersive audio. Photograph by Chris Michael reprinted with permission of Miller (2024). (b) Various artificial heads that represent a statistical distribution from three-dimensional scans of 350 adults. Photograph by Todd Welti.* ⏎

Figure 15.6(a) is a photograph of the lab showing a large spherical frame that holds 34 calibrated loudspeakers and four subwoofers on the floor below it. The setup can support playback of various immersive audio

formats including 22.2 channel, as well as measure free-field HRTFs. The loudspeakers facing the chair are on a 180° arc and spaced every 15°. With the listener sitting in the chair, small microphones are placed at the blocked entrance of their ears. The height of the chair is adjusted until their head is positioned in the laser calibrated position. For each sound source in the vertical array, an HRTF is measured and then the chair is rotated 15° to the right and the measurement process is repeated, until the chair has rotated full circle.

Statistical analysis of the HRTF dataset has identified the person closest to the average set referred to as ARCHIE (derived from *archetype*). Figure 15.6(b) is a photograph of several 3D-printed custom heads and pinnae. The four heads at the front represent the quintiles of the 350 humans in terms of the size and shapes of their heads and ears. Missing from the photograph is the average head ARCHIE. Each head can be equipped with different pinnae, microphones for measurements at the blocked entrance of the canal or an ear simulator for measurements at the DRP. The heads can be used for psychoacoustic experiments or to validate whether a headphone design has good reproducibility across a statistical distribution of heads and ears.

15.4.9 Measuring ANC Performance and Transparency in Headphones

Active noise cancellation (ANC) has become an important feature in consumers' headphone decisions. ANC can potentially reduce obtrusive background noise as much as 30 dB and greatly enhance the user experience (Rtings, 2024). ANC works by picking up the noise with a microphone outside the headphone (feed forward) or inside the headphone (feedback) and uses an out-of-phase version to cancel the noise. The hybrid ANC approach uses both feed-forward and feed-back microphones to further reduce the noise. The outside microphones can also provide a transparency or hear-through feature where the user can hear sounds around them with the convenience of not having to remove the headphones. Awareness of sounds around you is an important safety feature.

ANC is most effective below 1 kHz. Above 1 kHz, passive noise cancellation (PNC) does most of the work to provide a quieter listening experience. The mechanical design, shape and choice of materials for the cushions and ear tips are factors in how much insertion loss can be achieved. The filters in ANC headphones are traditionally fixed and tuned to cancel background noise produced by airplanes, which is relatively constant and predictable. Adaptive ANC can dynamically adjust the tuning of their filters and adapt to changes in the temporal and spectral characteristics of background noise. This makes them more effective across a wider range of use cases where you may be in an airplane, subway, restaurant, factory or driving a lawn mower.

IEC 60628-24 (2023) describes a method for measuring insertion loss from ANC and PNC using uncorrelated pink noise reproduced through eight loudspeakers to approximate a random incidence sound field at the headphone. The 1/3-octave spectrum of the noise is measured first with the headphones removed and then remeasured on the test fixture in both ANC on and off modes. The insertion loss provided by ANC, PNC and combined can be calculated and plotted as level (dB) vs. frequency, and the overall amount calculated as a single number in dB or phons.

Adaptive ANC measurements require a more ecologically valid test signal than uncorrelated pink noise, so engineers have created methods to record and reproduce different environmental soundscapes or noisescapes through an array of loudspeakers and measure how effective the headphone reduces the noise. A solution from Head Acoustics (Schäfer et al., 2019) uses an eight-microphone array to capture different noisescapes and reproduces the recordings through an 8-channel loudspeaker system. Butterworth (2024) visited Head Acoustics lab and describes the recordings ranging in level from 82 to 96 dBA. They include leaf blowers, lawnmowers, woodshop with power tools, airplane cabins and music venues. He provides a description of the measurements and some test results for two popular ANC headphones compared to two relatively inexpensive passive ear plugs (see Butterworth, 2024, figure 1). The entire process from measurement to processing results can take up to 3 hours but

yields several psychoacoustic metrics that capture how effective the ANC reduces noise in phons and improves the speech intelligibility.

SenseLab, a Danish acoustics research lab at Force Technology, has produced a large library of noisescapes captured with a high-order Ambisonics microphone array (SenseLab, 2024). The noisescapes are reproduced through a 26-loudspeaker playback system used to measure the objective and subjective performance of headphones in quiet and noisy environments. With the headphones placed on HATS binaural recordings are made with the headphones reproducing music and speech samples while different noisescapes (e.g. aircraft, factory) are reproduced through the loudspeakers as well as in silence. Recordings are made with and without the ANC mode engaged to determine how effective it works and its effect on the perceived sound quality of the signal.

A panel of consumers and trained listeners then evaluate the headphone recordings through a pair of calibrated headphones and provide ratings on different spectral and spatial attributes, as well as overall basic quality. The listening test results report significant differences in sound quality that vary depending on the performance of the ANC and type of noisescape played. The hear-through mode is also subjectively evaluated by comparing binaural recordings of speech and music reproduced through the loudspeakers with and without the headphones engaged in hear-through. The reference condition is the recording with the headphones removed from HATS.

As a paid service, SenseLab will provide audio companies comprehensive test reports that benchmark their products against competitors based on both objective and subjective measurements.

15.5 A HISTORY OF PREFERRED HEADPHONE CALIBRATIONS: FREE, DIFFUSE AND SEMI-REFLECTIVE FIELDS

Over the past 50 years the evolution of industry standards that recommend how to calibrate a headphone for best sound quality has been slow and

mostly uneventful. The free-field (FF) headphone calibration was popular in the 1970s but eventually replaced by the diffuse-field (DF) calibration standard in the 1980s. The DF calibration remains in standards today such as IEC 60268-7 (2010) and ITU-R BS. 708 (1990). The Harman research from 2012 to 2019 identified a target curve based on a loudspeaker in a semi-reflective field (SRF) that is preferred to both FF and current DF standards.

The following sections summarize the evolution of these different target curves including the rationale and research behind them.

15.5.1 Free-Field Calibration

In the 1970s headphones were calibrated by placing a loudspeaker directly in front of the listener in a reflection-free room. A tedious subjective loudness-matching procedure required the listener to adjust narrow bands of noise at different frequencies to match the loudness of the headphones to the FF by alternatively switching between headphone and FF conditions (with the headphone removed). This would be repeated for several test subjects to calculate the loudness transfer function that defined the headphone FF target curve. There are many potential reasons why this approach might not work for stereo.

Theile (1986) conducted formal listening tests and found the DF target to be preferred to the FF target that he claimed produced an unnatural timbre and in-head localization effects. The reason Theile argued is that natural hearing situations include both direct and reflected sounds from several directions. For the DF calibration sounds arrive equally from all directions, so there exists no directional or localization cues to conflict with our own FF HRTF. On the other hand, the FF headphone calibration is based on a single HRTF that most likely conflicts with our own and this highly directional cue will sound unnatural.

The FF target fell out of favor beginning in the 1980s, yet it remains part of the current headphone IEC 60268-7 (2010) standard today.

15.5.2 Diffuse-Field Calibration

Diffuse fields (DF) are produced when an omnidirectional sound source is placed in a reverberation room with little or no absorption, producing a random and equal distribution of sounds from all directions at the listener or microphone location. In the early 1980s a similar subjective loudness procedure as described above was employed but alternative methods were also explored. In one method, a probe microphone was placed in the ear canals of the listener to measure and then match the transfer function of the headphone to that of the sound field (Theile, 1986). A more repeatable and safer approach was to substitute the listener with a head and torso simulator (HATS). A third option was to use a headphone known to be DF calibrated and compare its performance with the headphone under test. More practical calibration methods were finally proposed.

Møller et al. (1995b) derived a headphone target curve based on different sound fields using a large set of HRTFs measured at the blocked ear canal. HRTFs define the transfer functions, both the frequency and phase responses at the entrance to the ear, for each direction and distance of a sound source. HRTFs capture both interaural time (ITD) and intensity (IID) differences and spectral cues that humans use to localize sound sources in space (Blauert, 1983). By selecting HRTFs from the appropriate directions and distances and integrating them, Møller et al. derived transfer functions of reference sound fields ranging from the FF to the DF and anything in between. This method eliminated the need for a physical reference sound field, making headphone calibration more practical and reproducible. A headphone could be measured and equalized to the DF target curve using a calibrated dummy head or ear simulator.

The DF headphone target was not seriously challenged until Lorho (2009) reported 80 listeners (25% audio engineers, 25% music students and 50% naïve listeners) preferred, on average, a significantly modified version of the DF target where its main feature, a wide 12 dB peak at 3 kHz, was reduced to just 3 dB. This paper sparked new interest to find better alternative headphone target curves to the ones recommended in the current headphone standards.

15.5.3 The Semi-reflective Field (SRF) Calibration

Typical consumer listening spaces are neither FF nor DF but rather something in between; a SRF where the loudspeaker produces a direct sound component followed by a few strong early reflections with most of the late reflections attenuated due to a combination of the directivity of the loudspeaker, propagation loss, air absorption, and the mid-high frequency absorption of reflecting surfaces in the room. When the steady-state in-room response of the loudspeaker is measured at the listening location it is dominated by off-axis (reflected) sound. Above the transition frequency it is well estimated by the predicted in-room response, a weighted combination of the listening window, early-reflections and sound power responses in a spinorama measurement (see Section 12.3.3). Low-frequency behavior is dominated by small room resonances (Chapter 14). Loudspeakers that are subjectively highly rated have flat and smooth on-axis/listening window (direct sound) responses (Section 2.7). Because virtually all loudspeakers are omnidirectional at low frequencies, becoming progressively more directional at higher frequencies, the result is a steady-state room curve that tilts downward by an amount depending on the directivity of the loudspeaker, the absorption characteristics of the room and the listening distance. Most importantly, the in-room measurement indicates the sound field is neither free-field nor diffuse.

Given that most stereo recordings are targeted for reproduction through loudspeakers in semi-reflective rooms, it makes sense such recordings will sound best when reproduced through headphones calibrated under these conditions.

Sank (1980) made similar proposals three decades earlier but never conducted formal listening tests that compared these targets with the DF target. Toole (1984) conducted informal listening tests on headphones that suggested that a calibration based a SRF might be a better choice.

Fleischmann et al. (2012) reported the first formal listening test results where three SRF headphone targets were evaluated. The targets were based on measurements of the steady-state in-room response of a 5.1-channel loudspeaker setup in an ITU-R BS.1116-3 (2015) specified room which is

more rigorously reflection controlled and acoustically damped than typical domestic rooms, placing the listener in a strong direct sound field above about 1 kHz. These were then equalized by three expert listeners to match the timbre of the speakers. Two of the SRF targets were found to be slightly preferred to the DF target, depending on the music programs. Other targets included the Lorho target, a flat target and three unequalized headphones that generally received lower ratings than the two SRF targets. Unfortunately, no measurements or details of the loudspeakers and the three SRF targets were given. The conclusions were that the SRF targets were equal to or better than the DF target, but the Lorho target was less preferred.

In summary, the FF and DF calibrations are based on sound fields that don't exist in semi-reflective rooms where audio content is produced and consumed. The FF calibration captures only the direct sound while the DF calibration captures both direct and reflected sounds but gives more importance to the reflected sounds. Both FF and DF calibrations fail to account for the directivity characteristics of typical loudspeakers, the room acoustics, and the reinforcement of bass due to solid boundary gain and room resonances. A SRF target curve accounts for these factors, and in theory should produce better sounding headphones. The Harman research discussed below was predicated on this hypothesis. But postulating a hypothesis was easy. Testing it involved a lot of hard work.

15.6 THE HARMAN RESEARCH (2012–2024)

In 2012, Olive and Welti began a 7-year investigation into the perception and measurement of headphone sound quality. Elizabeth McMullin joined the Harman research team and was succeeded by Omid Khonsaripour, both of whom made significant contributions. After 2019 the research effort was reduced, but it continues today.

15.6.1 Headphone Listening Test Methods: Physical Substitution

Conducting comparative blind listening tests on headphones presents several challenges not encountered when testing loudspeakers. There is no equivalent headphone version of the automated speaker shuffler that allows fast A/B/C/D blind comparisons that also removes positional biases from the test (see Figure 10.11). Instead, a test administrator must stand behind the listener and manually swap the headphones on demand. The listener must make fine adjustments to the positioning of the headphone to ensure it is properly seated and sealed. This interaction can inevitably lead to the identification of the headphone either by its weight, clamping force, comfort and feel of the cushions on the ears before the listener has even seen or heard it.

The first "blind" headphone listening test we conducted employed the physical substitution method (Olive et al., 2012), and we soon discovered it was a slow, tedious process for both the listener and test administrator. It also became clear based on feedback from the listeners that certain non-visual cues were influencing the results. The heaviest model (Audeze LCD-2) weighed 21 oz (595 g) and received the most consistent ratings. Listeners could easily identify it based on its weight alone. The test was effectively no longer blind.

There was a clear correlation between the perceived spectral balances and preference ratings and the acoustic measurements of the headphones. The most preferred headphones had the widest and smoothest measured frequency responses and were perceived to be spectrally well-balanced. But not all listeners agreed about the sound of a particular closed-back model that was perceived to have good bass for some listeners but lacking in bass for others. This headphone-listener interaction was found to be related to leakage and how well it fit and sealed on listeners' ears (see HP4 in Figure 15.8(d) in Olive and Welti, 2012). This set us on a path to investigate leakage and better control its effects.

15.6.2 The Virtual Headphone Method

The shortcomings of the physical substitution method led to consideration of two other methods:

1. Binaural recordings of the target headphones are made using a binaural manikin (or the listener's ears) and later reproduced through a calibrated reference headphone for evaluation (Hirvonen et al., 2003; Ravizza et al., 2023).
2. The reference or simulator headphone is equalized to simulate the measured transfer function of the different headphones being evaluated. We refer to this as the virtual headphone method. A similar approach was described by Briolle and Voinier (1992).

The binaural recording method has the advantage that the recordings include both the linear and nonlinear distortions of the headphones. In contrast, the virtual headphone approach includes only the linear distortions. However, the real-time equalization aspect of the virtual method offers greater flexibility in choosing test signals or equalizations during the test.

Both methods remove the influence of headphone visual and tactile biases and provide the listener immediate random access to the headphones under test. Both methods rely on the use of a replicator headphone that needs to be transparent and linear, so its performance does not influence the test results. A headphone with good reproducibility across multiple listeners is desirable so that the same signals are consistently delivered to their ears. In addition, controlling leakage in the replicator headphone is important.

The virtual headphone was ultimately our choice for several reasons:

- The non-linear distortions in nearly all headphones we tested were not sufficiently audible to influence the test results.
- We were able to closely match the frequency response of the target headphone using the replicator headphone because of the minimum-phase behavior of the headphones.
- Validation tests of the virtual headphone method produced similar headphone rankings as using the (1) actual headphones (Olive et al.,

2013c) or (2) binaural recordings of the headphones (Olive et al., 2016b).

The virtual method produced preference ratings of AE headphones that were highly correlated (Pearson correlation coefficient $r = 0.86$) with results using the real headphones via the physical substitution method (Olive et al., 2013c). Perfect correlations are not likely possible because the leakage effects and tactile-weight-related biases observed in tests using real headphones are not present in the virtual headphone tests.

Briolle and Voinier (1992) reported sound quality evaluations of 12 headphones made by 22 listeners using both real and virtualized versions. The correlation between the two sets of results was 0.93. The virtual method reduced the time switching time between two headphones from 5 s to 10 s to about 0.5 s, and it improved the consistency in listeners' ratings. Similar results were produced using three different qualities of replicator headphones.

When IE headphones were evaluated using the virtual method the headphone preference correlations with results from binaural recording methods were even higher: $r = 0.95$ and 0.99 (Olive et al., 2016b). Leakage effects were carefully controlled for both measurements and listening tests putting a small MEMS microphone inside the ear tip the replicator IE headphone to monitor leakage. The agreement in results between virtual and binaural recording methods suggests that any nonlinear or non-minimum phase behavior in the headphones present in the binaural recordings (but not in the virtual versions) had little effect on preference. However, in cases where the nonlinear distortion is well above audible threshold, the binaural recordings can reveal its negative effect on listeners' preference ratings (Temme et al., 2014).

15.6.3 The Harman Headphone Target Curve

The origins of the Harman headphone target curve are based on an accurate professional 7.1 monitor system (JBL LSR6332, but using only the left and right channels) designed to produce flat direct sound and smooth sound

power in a semi-reflective listening room (Olive et al., 2013b). The steady-state in-room response at the listening position closely matches the predicted-in-room (PIR) of the loudspeaker. A similar result was found using a stereo pair of Revel Performa F208 loudspeakers (Olive and Welti, 2015) that also produces neutral direct sound, smooth sound power and directivity. This can be verified by examining their comprehensive anechoic spinorama measurements (see Figure 17.2 at $5000/pair).

The dimensions of the room are rectangular: 21 ft (6.4 m) × 24 ft (7.32 m) × 9 ft (2.74 m), and its dimensions and reverberation time (0.4 s) are typical of domestic listening spaces. More details of the room with acoustic measurements can be found in Olive (2009).

Figure 15.7(a) shows the physical set up of the loudspeakers and the position of the acoustic test manikin (GRAS Type 3.3) located in the primary listening seat. The steady-state frequency response of the left and right loudspeaker was measured separately with the manikin positioned at 0° and ±30° and smoothed in the frequency domain. All the measurements shown here are based on an average of 6 measurements (2 ears × 3 angles) with 1/3-octave smoothing. Figure 15.7(b) shows the response of the headphone fixture after the loudspeaker has been flattened. This is the baseline Harman target that listeners later adjusted the bass and treble based on preference. The curve labeled “After EQ” is an example of what the target curve would be if their adjustments produced the EQ filter shown here. Figure 15.7(c) shows results from two separate listening experiments where listeners adjusted bass and treble filters to produce two slightly different Harman targets. In the first experiment (Olive, 2013) listeners adjusted the bass from flat to +5 dB and the treble from flat to −3 dB. In the later experiment (Olive and Welti, 2015), listeners slightly increased the bass to +6.4 dB and increased the treble to −2 dB. Details on these experiments are discussed in later sections. Figure 15.7(d) shows what the flattened in-room loudspeaker measurement looks like from the perspective of the microphone at the primary seat with the headphone test fixture removed. The after-EQ curve is the preferred headphone EQ applied to the loudspeaker. It closely resembles the PIR of the Revel F208 loudspeaker

shown here, which is approximately 10 dB downward sloped curve from 20 Hz to 20 kHz. The PIR is approximately how the loudspeaker measures in the room (ignoring room resonances) before flattening EQ is applied.

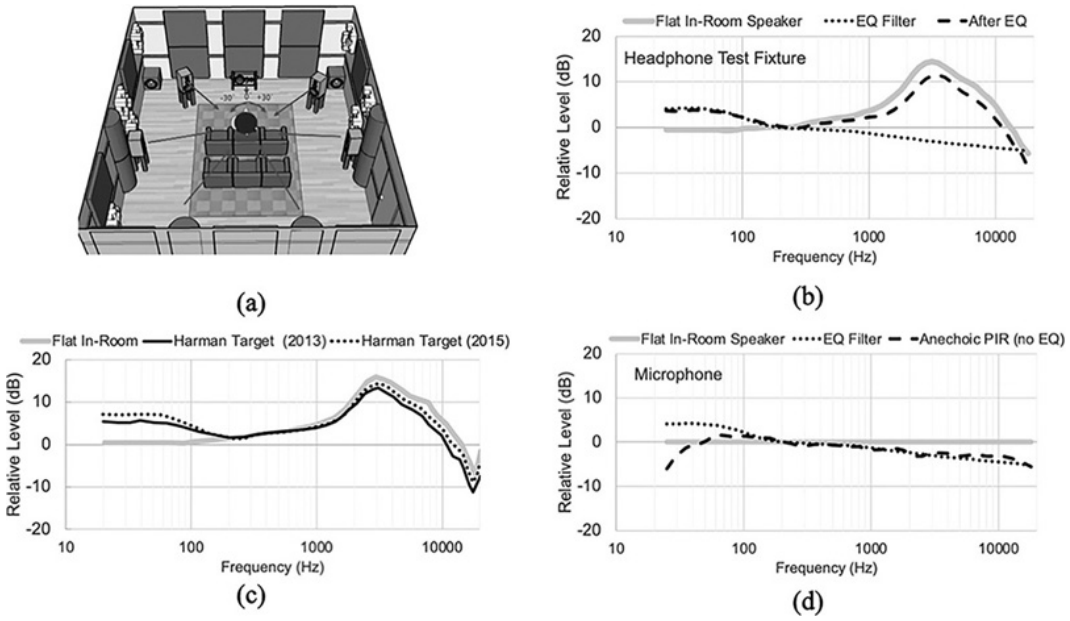


FIGURE 15.7 *All the measurements shown have been 1/3 octave smoothed. (a) The Harman target curve for AE/OE headphones began by measuring the steady-state response of a neutral JBL professional monitor and later a Revel loudspeaker in the Harman Reference Room and equalizing it to be flat. (b) The result measured with a GRAS Type 3.3 ear coupler and custom head produces the thick gray curve. Using a method of adjustment listeners adjusted the bass and treble levels of the headphone according to preference. After the EQ (dotted curve) is applied it produces the target curve (dashed curve). (c) The solid black and dotted curves are preferred AE/OE headphone target curves from two experiments: Olive et al. (2013a) and Olive and Welti (2015) using a method of adjustment. The 2015 target has slightly more bass and treble as it included both trained and untrained listeners. (d) After EQ (dotted curve), the headphone target approximates the predicted in-room response of the Revel Performa F208 (dash curve) before bass shelf is applied.* ⏎

In summary, the Harman headphone target curve closely approximates the PIR response of a well-designed loudspeaker, like the Revel F208. If we flatten the in-room response of such a loudspeaker it becomes the equivalent of the DF headphone target curve or a loudspeaker with flat sound power. When you give listeners bass and treble controls to adjust this target, they will restore it back to its original response adding more bass and reducing the treble, whether listening to it through a loudspeaker or a headphone. The finding that there is a preferred target curve common to both loudspeakers and headphones, while not surprising, is encouraging.

With a common target, a recording made to sound good through loudspeakers will also sound good through headphones. And when consumers play back the recordings through their loudspeakers and headphones, they too will hear good sound and what was intended. This is what recording mixers refer to as "good translation" (see Section 12.5). The recording will sound similar or translate from loudspeaker to headphone, and from studio to the home. A common target between loudspeakers and headphones goes a long way to solving Audio's Circle of Confusion. But first, more evidence is needed to determine if a SRF target curve like the Harman target fits the bill.

15.6.4 Testing Harman Target Curve Against Three Highly Regarded Headphones

The first version of the Harman target curve (Olive et al., 2013a) was tested against other targets using the virtual headphone method described above. Listeners strongly preferred the Harman target in descending order of preference to two DF targets (Møller et al., 1995b), two high-quality headphones, the Lorho target, and the FF target. The trained listeners described both DF targets as having too much emphasis in the upper midrange (2–4 kHz) and lacking bass. The Lorho target had too little energy at 2–4 kHz, which made instruments sound "muffled and dull". The FF target was strongly criticized for its strong emphasis between 2 and 4 kHz, lack of bass, and harsh and nasal colorations. Listeners described the highest rated Harman target as having "good bass with an even spectral

balance". The measured frequency responses of the headphone targets correlated to and confirm listeners' descriptions of their sound quality (see Olive et al., 2013a, figure 2). (See OR Chapter 15 for more details.)

The initial Harman target curve test results used a small sample of ten trained listeners. While the results were encouraging, there remained many questions to address. To better understand if certain demographic factors might influence the acceptance of the curve, it was tested using a larger number of listeners from a broad range of ages, listening experiences, and geographic regions. The target curve was benchmarked against three headphones considered industry references at the time in terms of sound quality or commercial sales (Olive et al., 2014). They ranged in price from $269 to $1500 and included dynamic and magnetic planar transducer designs. A total of 283 listeners participated from four different countries (Canada, United States, Germany and China) and included a broad range of ages, listening experiences and genders. Most of the participants were Harman employees.

Figure 15.8 shows the mean headphone preferences across different locations for (a) trained and (b) untrained listeners, (c) different countries and (d) different age groups. The results show that headphone preferences were remarkably consistent across the 11 test locations for both trained and untrained listeners. As expected, the (a) trained listeners were more discriminating and consistent than the (b) untrained listeners. Headphone preferences were also relatively consistent across the (c) different countries and (d) different age groups. The exception was listeners in the 55+ age category who tended to prefer HP2, a brighter headphone with less bass than the Harman target curve. A possible explanation could be age-related hearing loss; additional treble and less bass can help improve intelligibility. More research is needed to provide definitive answers.

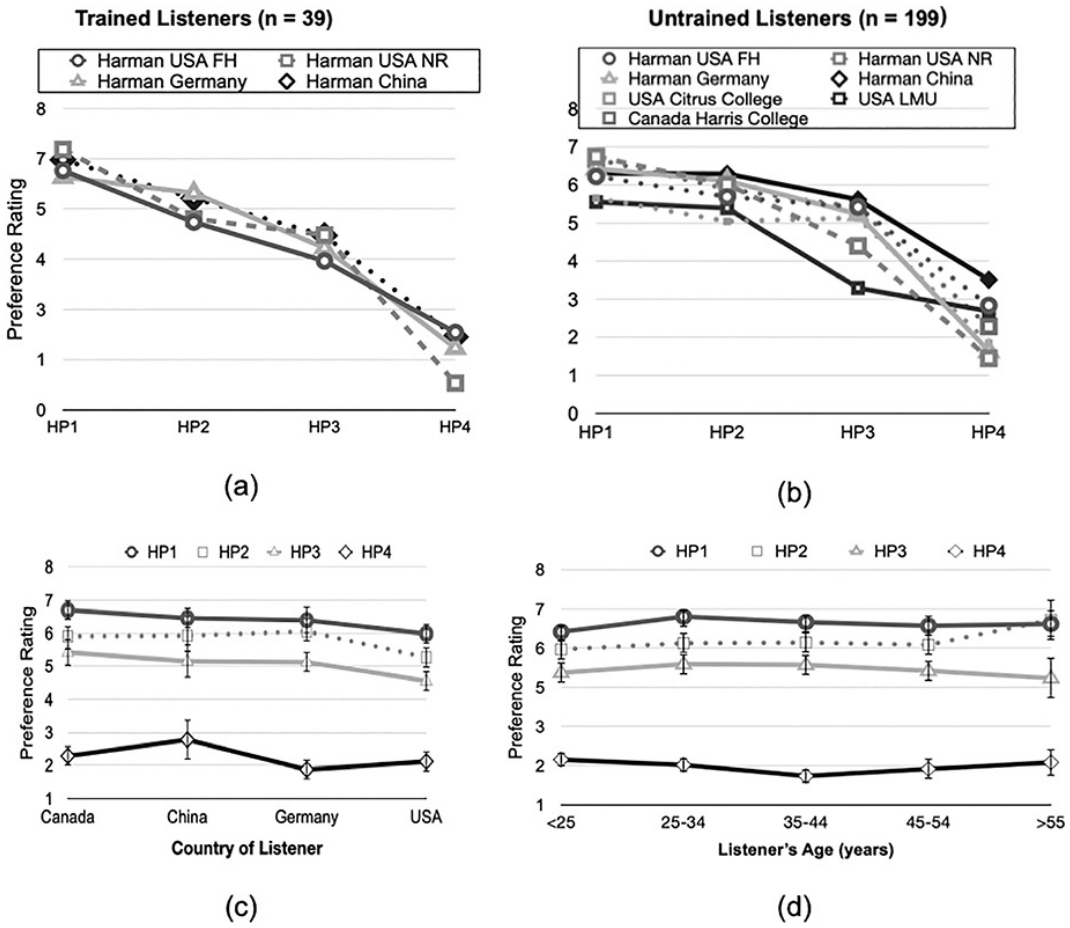


FIGURE 15.8 *(a) The mean preference ratings are shown for 11 different groups of listeners categorized as four groups of trained) and (b) seven groups of untrained (right). (c) The mean preference ratings and 95% confidence intervals for the four headphones for listeners tested in four countries and (d) within different age groups. HP1 is the Harman target curve and HP2 and HP3 are high-quality, high-priced headphones. HP4 was the most popular headphone in terms of sales (Olive et al., 2014).* ⏎

15.6.5 Listeners' Preferred Levels of Bass and Treble

The same listeners participated in a second experiment where they adjusted the bass and treble levels of the headphone several times according to taste using three different samples of music, Olive and Welti (2015). Figure 15.9 plots the mean preferred bass and treble levels relative to the baseline flat-

in-room Harman target curve (thick gray curve in Figure 15.7(b)). On average listeners adjusted the bass 6.4 dB and the treble 2 dB down relative to the flat baseline target, the gray curve in Figure 15.7(b). They increased the bass +1.5 dB and the treble 1 dB relative to the previous adjustment experiments (Olive et al., 2013a).

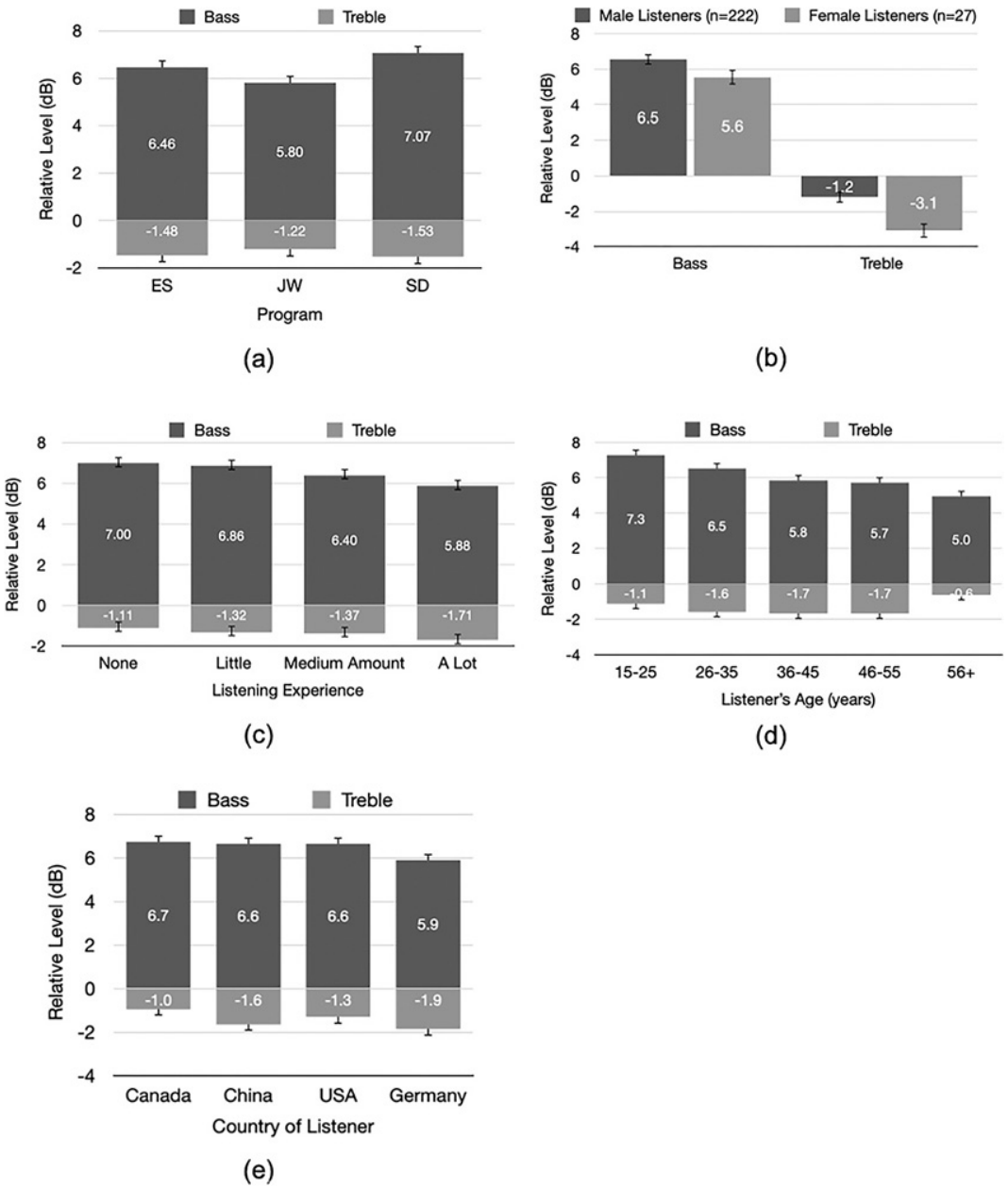


FIGURE 15.9 *The mean bass and treble levels and 95% confidence intervals using a method of adjustment. Listeners adjusted the levels of the headphone after it was calibrated to a*

flat steady-state response of a loudspeaker measured in a reference listening room. The results are shown for (a) program, (b) gender, (c) listening experience, (d) age and (e) the country of the test location (Olive and Welti, 2015). ⏎

The preferred adjusted levels were influenced by several factors: the (a) music program (b) the listener's gender, (c) age, and (d) prior listening experience. The program interactions with preferred bass and treble levels can be explained in part due to circle of confusion issues that cause variability in the quality of recordings (Section 1.7).

Female listeners preferred on average about 1 dB less bass and 2 dB less treble than their male counterparts. Younger and less experienced listeners preferred more bass and treble than their older, more experienced counterparts. The older listeners (55+ years) were the exception here, preferring significantly more treble and less bass, consistent with their preference for headphone HP2 that has less bass and more treble than the other three models. Altogether, these findings suggest that a single headphone target may not be sufficient to satisfy variations in the recordings, individual tastes, listening experience, and hearing loss. A simple solution is to provide simple bass and treble controls that allow listeners to compensate for these variations.

15.6.6 Benchmarking the Harman Target Against 31 Models

The previous section showed evidence that the Harman target was well accepted by a large sample of listeners from four different countries across a range of ages, gender and listening experiences. However, the results were based only on four headphones. A new experiment was conducted to address this shortcoming.

A total of 31 different headphone models from 18 manufacturers were evaluated by 130 listeners, with an approximately equal number trained and untrained (Olive et al., 2018a). The headphones ranged in price from $60 to $4000, including open- and closed-back designs with dynamic or magnetic planar drivers. The same virtual headphone double-blind method was used

to eliminate biases from visual and tactile cues. The results established that, on average, both trained and untrained listeners preferred the headphone equalized to the Harman target in 28 of the total models tested. Four models with frequency responses close to the Harman target were equally preferred (see OR Chapter 15 for more details).

15.6.7 Segmentation of Listeners Based on Preferred Headphone Sound Profile

The results from the above study described were further explored using a statistical method known as agglomerative hierarchical clustering (Olive et al., 2019). The purpose was to determine if there existed different segments or classes of listeners based on similarities in their headphone preferences, what their preferred sound profiles were, and if there were any underlying demographic factors that might predict class membership. There was already evidence that younger males and less experienced listeners preferred higher levels of bass and treble in their headphones compared to females, experienced, and older listeners (Olive et al., 2013b; Olive and Welti, 2015). A reasonable hypothesis was that segmentation of headphone preferences may relate to preferred bass and treble levels, and certain demographic factors might be associated with the preferred sound profile.

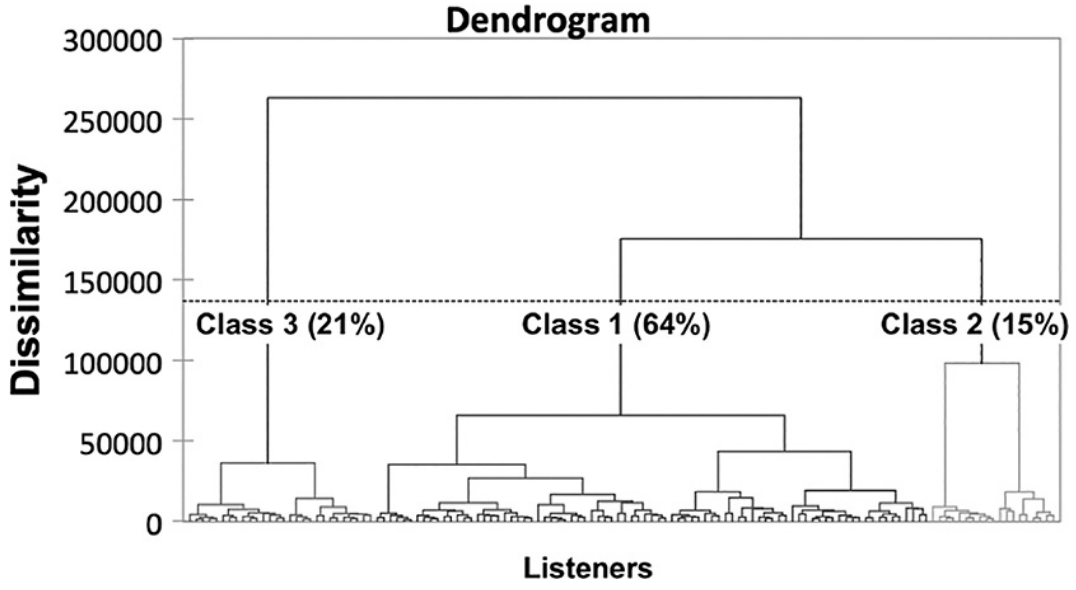


FIGURE 15.10 *A dendrogram based on Agglomerate Hierarchical Cluster (AHC) analysis of 130 listeners who gave preference ratings for 32 different models of AE headphone and one equalized to the Harman target curve. Listeners are assigned to one of three clusters based on (dis)similarities in their headphone preference ratings with the percentage of total listeners in each class shown.*

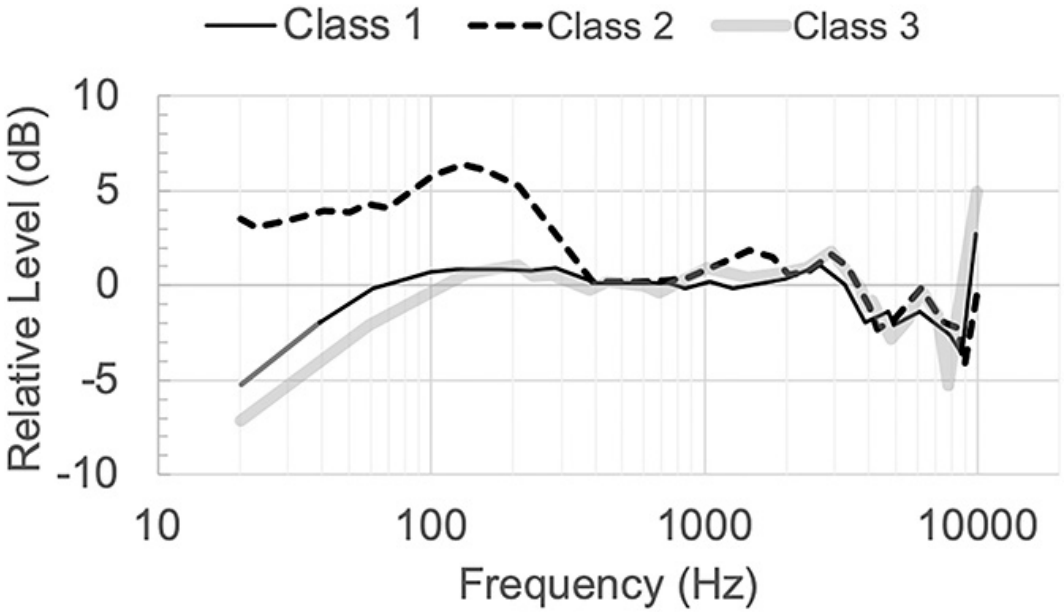


FIGURE 15.11 *The preferred AE/OE headphone. sound profiles of the three classes of listeners identified from segmentation analysis. Each curve is based on the average of the five highest rated headphones in each class relative to the Harman target curve. Class 1 prefers headphones that on average closely comply with the Harman target. Class 2 prefer about 4 dB more bass than the Harman target curve and Class 3 prefers 2–4 dB less bass.*

A dendrogram of the cluster analysis reveals three clusters or classes of listeners based on (dis)similarities in how they rated the 32 headphones (see Figure 15.10). Class 1 accounts for 64% of the listeners, with two smaller

classes accounting for 15% (Class 2) and 21% (Class 3) of the total listeners tested.

The preferred sound profiles associated with each class were estimated by calculating the average frequency response of the top five rated headphones in each class. The sound profiles are shown in Figure 15.11 relative to the Harman target curve. Class 1 prefer headphones that on average mostly comply with the Harman target. Class 2 prefer the Harman target with 4–6 dB more bass. Class 3 prefer the Harman target with about 2 dB less bass. There is also some indication less treble between 4 and 9 kHz may be preferred

Figure 15.12 summarizes the three different classes of listeners, their sound profiles and the demographic factors associated with each class. Sixty-nine percent of the males in the study are members of Class 1 (*Harman Target Lovers*) compared with 56% of females. Class 1 has roughly equal representation from trained (70%) and untrained (65%) listeners. Class 2 (*More Bass Is Better*) has the fewest members overall and is represented by all categories except female; only 4% of females tested want more bass than the Harman target provides. Class 3 (*Less Bass Is Better*) members are disproportionately represented by females (40% of females are in this class vs. 13% of males) and listeners over the age of 50 (50%). Hearing loss may be a confounding factor here. More research is needed to better understand the role it plays in headphone sound quality preferences.

Class 1: (64% of listeners)
Demographic Profile: Includes all categories (gender, trained/untrained, ages except fewer over 50 years old).
Sound Profile: Prefer headphones tuned (or close to) Harman target curve

Class 2: (15% of listeners)
Demographic Profile: Male, younger listeners (<50 years).
Sound Profile: Prefer headphones with 3-6 dB more bass than Harman target below 300 Hz, and +1 dB above 1 kHz.

Class 3: (21% of listeners.)
Demographic Profile: Untrained, disproportionate percentage of females, and people 50+ years old.
Sound Profile: Prefer 2-3 dB less bass than Harman target and +1 dB more treble above 1 kHz.

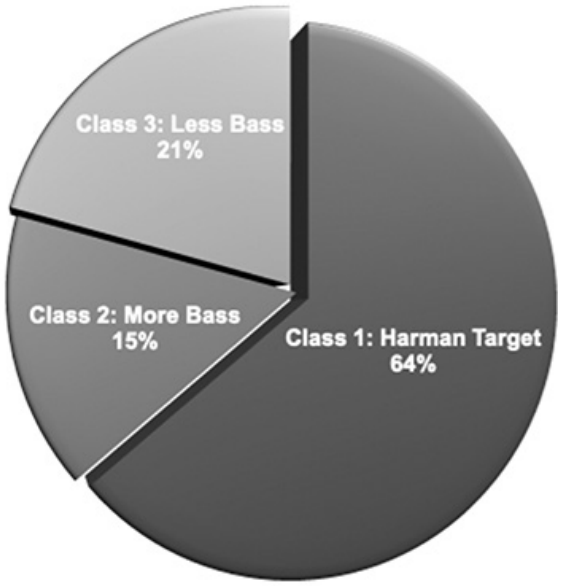


FIGURE 15.12 *The three classes of listeners based on segmentation analysis where they are clustered according to dissimilarities in headphone preference ratings. The percentage of listeners in each class is shown along with their demographic profile and preferred sound profile relative to the Harman target curve (Olive et al., 2019).* ⏎

The main takeaway is that the Harman target is a good starting point for headphone design because it satisfies the tastes of most listeners (64%) over a broad range of age groups, genders, and levels of listening experience. The two smaller classes of listeners who prefer headphones with more bass or less bass can be accommodated through a simple bass tone control on the headphone or via an app on the playback device. The bass adjustment would also help compensate for inconsistences in the quality of recordings that contain either too much or not enough bass and treble. A word of caution: the research suggests adding too much bass beyond the Harman target may alienate many listeners given that the "more bass is better" segment is a small segment with little female and older listener membership. Conversely, reducing the bass too much may alienate trained

and experienced listeners who are underrepresented in the “less bass is better” segment.

15.6.8 Predicting Listener Headphone Preferences Based on Objective Measurements

Conducting controlled headphone listening tests is a challenging, time-consuming and expensive proposition. Section 11.1 described a statistical model to predict listener loudspeaker preference ratings based on objective measurements that reduces—not eliminates—the reliance on listening tests (Olive, 2004). Loudspeakers that deviated from a flat, wide and smooth frequency response were penalized with lower preference ratings. A similar approach was taken here to predict listener’s headphone preference ratings using objective measurements (Olive et al., 2018a).

The 31 headphones from the same study discussed by Olive et al. (2018a) were sorted into four categories of sound quality based on listeners’ preference scores: excellent (90%–100% preference rating), good (65%–76%), fair (42%–54%), and poor (0%–39%). Frequency response measurements of the headphones were made using the GRAS 45CA Type 3.3 ear simulator equipped with a custom Welti (2015) pinna.

Figure 15.13 plots average frequency response for the headphones within each sound quality category along with the Harman target curve. The difference between the two curves is referred to as the error curve. The error curve is shown at the top of each graph (dashed curve) along with a regression line that best fits the error curve. Observing these four graphs, the relationship between objective and subjective headphone measurements is visually quite clear: the more the frequency response of a headphone deviates from the Harman target curve, the lower the listeners rated its sound quality. The further the slope of the error curve deviates from a flat line at 0 dB, the lower the preference rating. The deviations could be caused by resonances or bandwidth limitations but not specifically weighted according to frequency or Q. The slope of the error curve is a good indicator of the spectral balance with large values indicating the headphone could be perceived as too bright/dull or too bassy or thin.

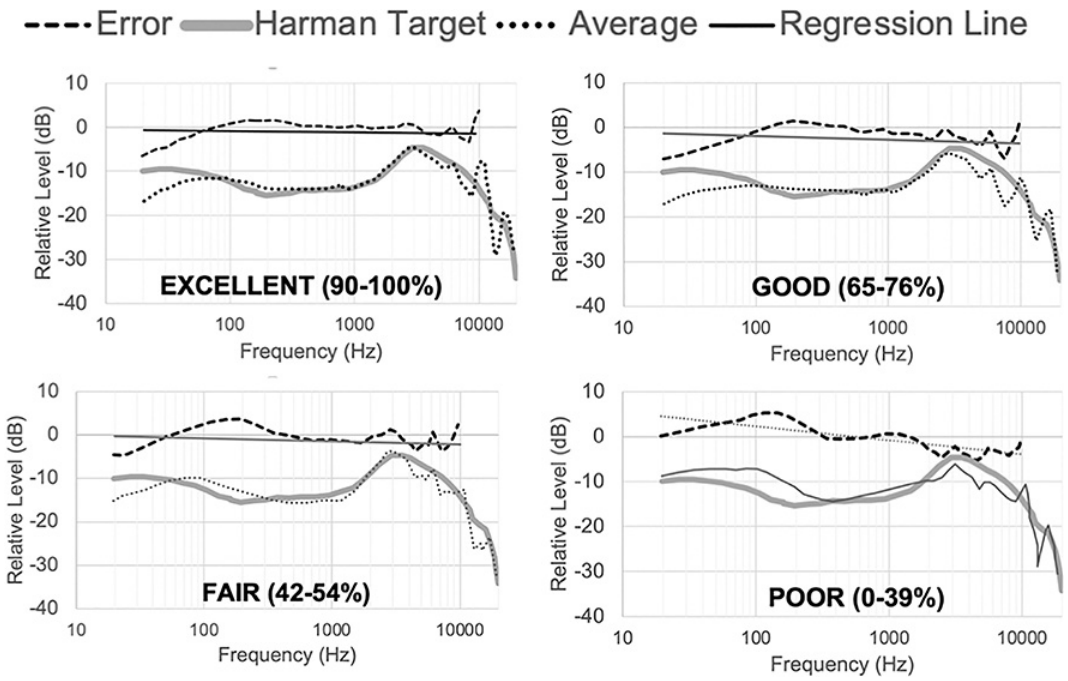


FIGURE 15.13 *The average frequency response (dotted) and error curves (dashed) for 31 around the ear (AE) headphones assigned to four categories of sound quality based on their preference ratings given in controlled listening tests (Olive et al., 2018a). The Harman target curve (solid gray) is used to calculate the error curve. The regression line (thin solid line) best fits the error curve.*

A linear model was developed using partial least squares (PLS) regression and two variables:

SD: The standard deviation of the error defined by the headphone error curve calculated from the *y*-values from 50 Hz to 10 kHz as defined by Equation 15.1:

$$SD = \sqrt{\frac{\sum (y-y)^2}{(n-1)}} \qquad \text{Equation 15.1}$$

AS: The absolute slope (*AS*) of a logarithmic regression line that best fits the error curve from 50 Hz to 10 kHz according to Equation 15.2.

$$AS = \left| \sqrt{\frac{\sum \ln(x) - \ln(\bar{x})(y - \bar{y})}{\sum (\ln(x) - \ln(\bar{x})^2)}} \right| \qquad \text{Equation 15.2}$$

Equation 15.3 describes a linear model that predicts the preference rating (*PPR*) of the headphone using the two variables *SD* and *AS*.

$$\boldsymbol{PPR} = 114.49 - (12.6 \times \boldsymbol{SD}) - (15.52 \times \boldsymbol{AS}) \qquad \text{Equation 15.3}$$

Including frequencies below 50 Hz and above 10 kHz in the calculation of the variables did not further explain the underlying variance in headphone preference ratings. The standardized coefficients for the two variables in the model are weighted approximately equal: *SD* = −0.47 and *AS* = −0.434. Both are equally important in predicting headphone preference.

Figure 15.14 shows a scatterplot of the predicted vs. measured preference ratings for the 31 headphones with the 95% confidence intervals. The Pearson correlation coefficient ($r = 0.86$) indicates good agreement between the predicted vs. measured preference ratings with a root mean square error (RMSE) of 6.7 preference rating on a 100-point scale. Two of the headphones (HP6 and HP30) were outliers and fall outside the confidence limits: the predicted ratings were much higher than the ratings they received. The measurements of the headphones suggest the lower observed ratings could be attributed to a large resonance above 1 kHz that was underestimated given by the fact that the explanatory variables *SD* and *AS* don't consider the Q or frequency of the deviation, only their magnitude (see Olive et al. (2018a, appendix 2)). A future version might give more weight to deviations associated with resonances. If HP6 and HP30 are excluded from the model, the goodness of fit improves to $r = 0.90$.

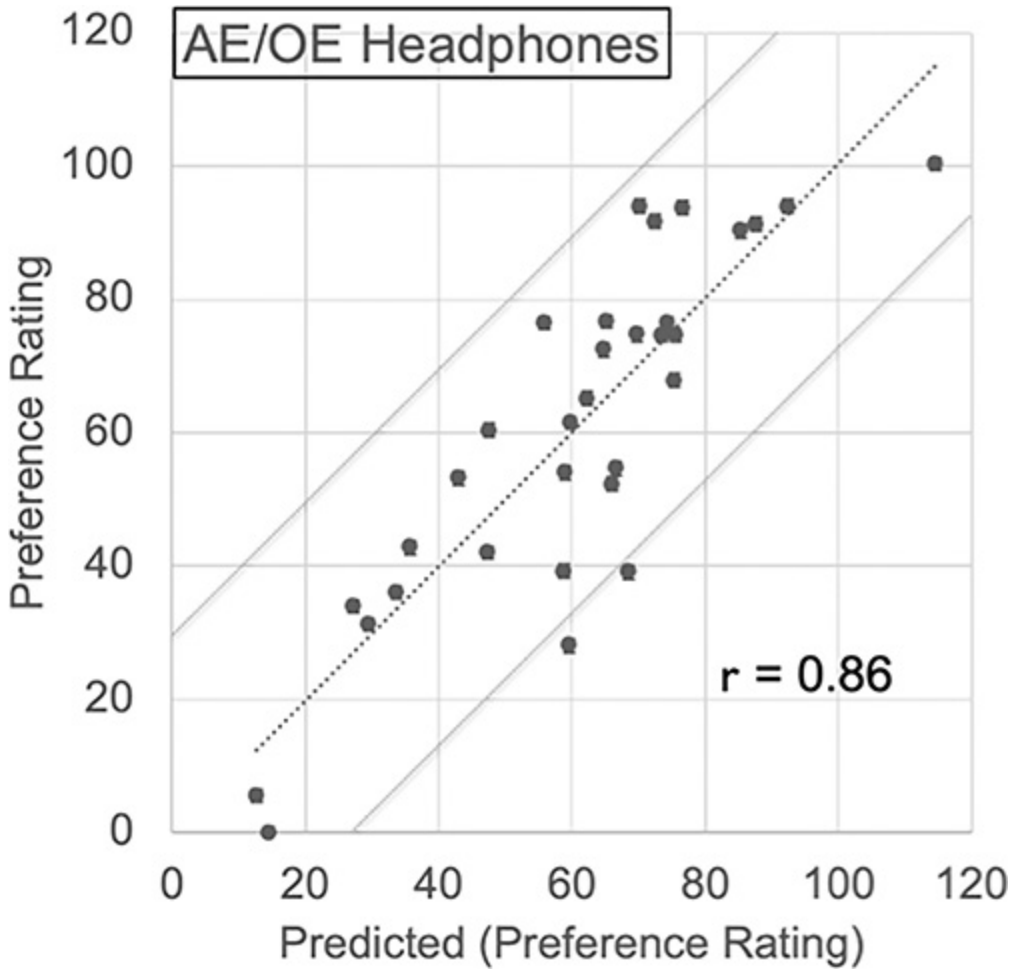


FIGURE 15.14 *The predicted vs. observed preference ratings for the 31 different AE/OE headphones based on the multiple regression model described in Olive et al. (2018a). Correlation: r = 0.86.* ⏎

A similar model was developed for IE headphones based on evaluations of 31 IE headphones made by 71 listeners (Olive et al., 2017b). It produced even higher correlations (r = 0.91) between predicted and observed preference ratings with a smaller error: RMSE = 5.5 preference rating. IE headphones tend to be more well behaved with smooth responses fewer resonances at higher frequencies, whereas larger diaphragm AE/OE headphones tend to misbehave. This could explain why the IE model is slightly better at predicting listener preference ratings. (See OR Chapter 15.)

15.6.9 Harman IE Headphone Target for B&K 5128

The Type 4.3 ear simulator introduced in 2018 with the B&K 5128 accurately simulates the acoustic impedance from 20 Hz to 20 kHz of an average human ear (see Section 15.4.2). This makes it a good choice for designing and testing headphones since the measurements will more accurately estimate how the headphone performs on humans. The Harman target curve was developed on the older Type 3.3 ear simulator which is only accurate from 100 Hz to 10 kHz.

A simple correction that translates measurements from Type 3.3 to Type 4.3 is complicated by the fact that the acoustic input impedances of the ear simulators are different over much of their frequency range. This means that the differences in frequency response will vary depending on the acoustic load of the headphone and how it interacts with the acoustic impedance of the ear simulator (Mansilla, 2020; Johansen et al., 2024). Since IE headphones generally have higher acoustic output impedances their frequency response will be most affected by the acoustic input impedance of the ear simulator.

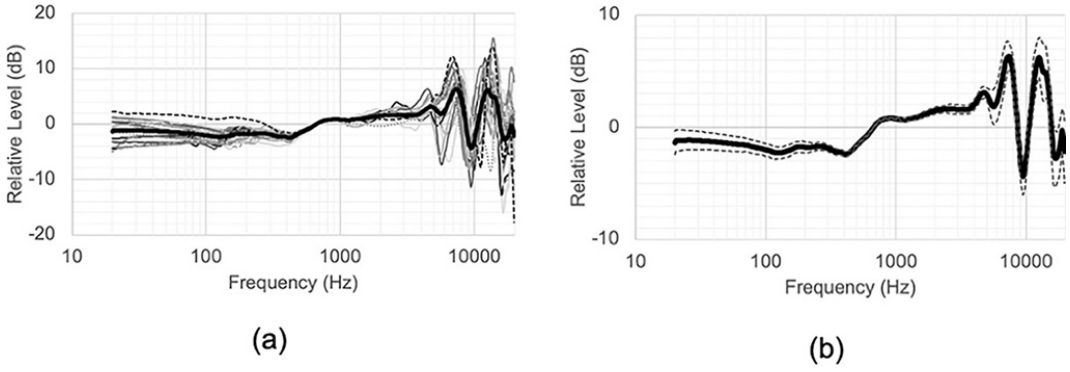


FIGURE 15.15 *(a) The difference in frequency response of 20 models of IE headphones measured on a Type 3.3 ear simulator (GRAS RA0045) and Type 4.3 ear simulator (B&K Type 4620). The dark line is the average of the 20 headphones. (b) The average difference is plotted alone with the 95% confidence intervals (note the expanded vertical scale).*

Figure 15.15(a) illustrates the interaction by plotting the difference in frequency response of 20 different models of IE headphones measured on both Type 4.3 and Type 3.3 ear simulators (Olive, 2024). The mean response of the sample is shown as the thick black curve. The difference in magnitude varies between models with 10 dB differences below 100 Hz and more than 20 dB differences above 10 kHz. The average difference with 95% confidence is replotted in Figure 15.15(b). On average a headphone measured on the Type 4.3 will have about 2 dB less bass below 500 Hz and 2 dB more treble above 2 kHz with two ±5 dB narrow peaks and dips above 7 kHz compared to how it measures on the Type 3.3.

The conclusion from this exercise is that a single universal correction/compensation between ear simulators is not possible since the magnitude of the error is headphone model dependent. However, it is valid to calculate a correction for a single headphone model, for example, the IE headphone used for the Harman target curve research. Using this procedure, we translated the Harman IE target from the Type 3.3. to the 4.3 ear simulator. The target curve was then validated by comparing its performance in controlled listening where 32 trained listeners compared it to four other targets using a virtual headphone test method. Figure 15.16 shows the frequency responses targets measured on the B&K 5128. The five IE headphone target curves were:

- *SoundGuys*—The target curve discussed in Section 15.8.3.
- *Harman 2019*—The Harman IE target curve (Olive et al., 2017a) translated to the B&K 5128 as discussed above.
- *DF*—The diffused field response of the B&K 5128 (provided by B&K).

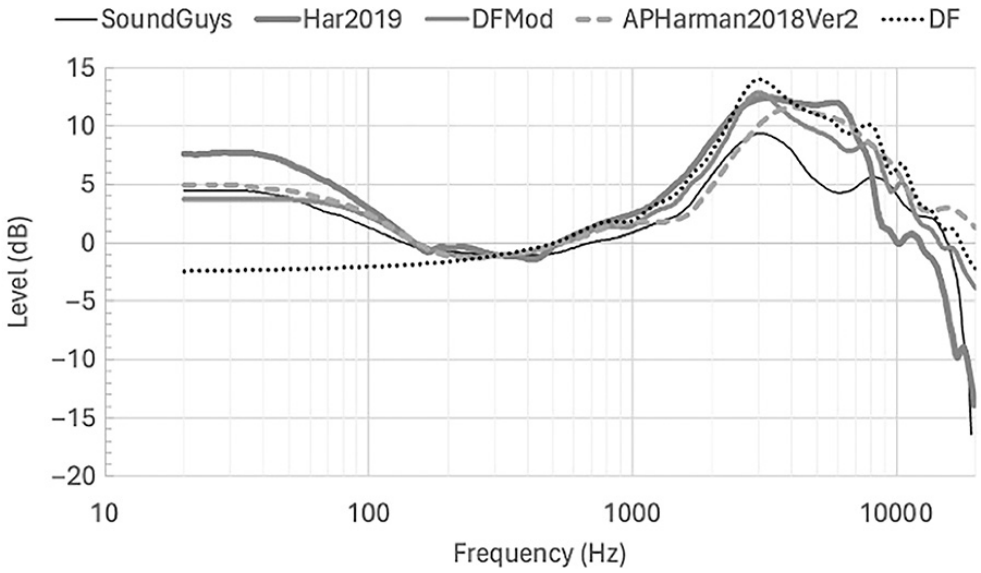


FIGURE 15.16 *The five different IE target curves evaluated by 32 trained listeners in this study measured on the B&K 5128 test fixture (Olive, 2024).*

- *DFMod*—The DF response modified to have more bass and less treble.
- *APHarman2018Ver 2*—The highest rated curve in the SenseLab study (Section 15.8.5).

Figure 15.17 displays boxplots showing the distribution of preference ratings for each target curve based on evaluations from 32 listeners. The boxplot graphically shows the estimated mean, median, and upper and lower quartiles of all ratings with the vertical lines or whiskers indicating ratings beyond the quartiles, and the single dots indicate outliers. Both the *SoundGuys* and the *Harman 2019* targets were equally preferred, and both were preferred to the other target curves. The lowest rated target was the *DF*, which is consistent with results from previous studies reported in this chapter. *DF Mod* which adds bass and reduces the treble of the *DF* improves its sound quality. The modified version of the Harman 2018 Target (*APHar2018Ver2)* that received the highest ratings in the SenseLab study did not fare as well in this study. This could be related to differences in the test methodology.

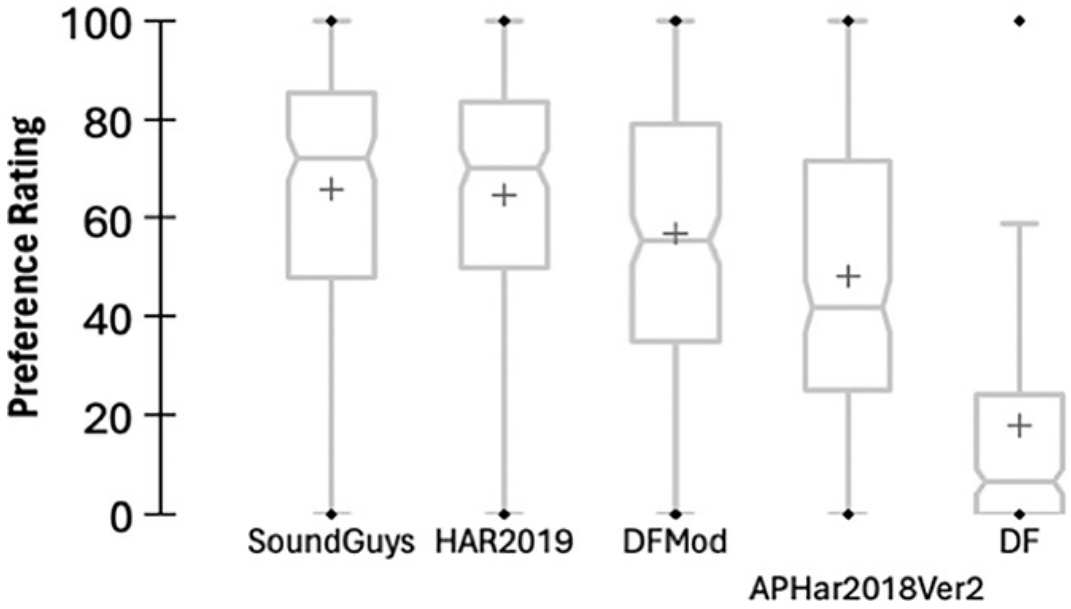


FIGURE 15.17 *Boxplots of the distribution of 32 listeners' preference ratings for the five target curves showing the mean (+), the median, and the upper and lower quartiles. The whiskers and dots indicate the range of values outside the quartiles.* ⏎

Segmentation analysis revealed there were two clusters of listeners based on preference. The largest cluster (73% of listeners) equally preferred the *SoundGuys* and *Harman 2019* target curves. The smaller cluster (27%) preferred *DFMod* slightly to *APHar2018Ver2*, both being brighter targets with more treble. Listeners in this group were on average 10 years older than the larger cluster. More details are in OR Chapter 15.

In summary, the *Harman 2019* IE target as defined on the B&K 5128 was rated highest in the blind listening and equally preferred to the *SoundGuys* target by 73% of the listeners. Five years later since the first validation study (Olive et al., 2017a) of the Harman IE target, the translated B&K 5128 version of it continues to hold its own.

15.6.10 Proper Application of the Harman Research

The Harman headphone research has produced over 20 scientific papers since 2012 that have received wide attention and scrutiny within the scientific and headphone communities. There has been considerable

coverage in headphone review organizations including Innerfidelity, Audio Science Review, SoundGuys, Head-fi.com, Crutchfield.com, Headphones.com and others. The Harman target is often included as a benchmark in headphone reviews that include measurements because it has been rigorously validated with controlled blind listening tests. Enthusiast sites like www.autoeq.app allow users to equalize their headphones to the Harman target or predict their headphone's preference rating using the Harman model. While the enthusiasm for science is encouraging there has been a fair amount of misunderstanding and misapplication of the research: this can lead to disappointment and false expectations about what the research can and cannot do. Here are some of the caveats and suggestions in helping the proper application of the research and interpretation of the results:

- Headphones can have similar predicted preference scores and sound quite different since the model treats deviations from the target equally regardless of their frequency and Q.
- In the interpretation the error curve and its slope keep in mind that broadband (low Q) deviations and spectral tilts are more audible compared to narrow-band high Q deviations that produce a flatter slope.
- The error of the predicted rating is 6.7 rating on a 100-point scale. The differences between two headphones with scores within this rating range are not statistically significant.
- The model that predicts the preference rating does not consider nonlinear distortion, interchannel level or phase difference. Other important factors such as comfort and ANC performance are not considered.
- Leakage effects in both objective and subjective measurements were controlled throughout the Harman research. In practice, if the headphone measurements don't reflect the leakage effects experienced

on humans the objective measurements and predicted ratings will poorly correlate the subjective impressions.

- The predicted preference rating is a statistical average of all listeners tested. It does not account for your personal tastes, listening experience, hearing loss, and the kinds of recordings you listen to.
- The model is based on headphones measured on a GRAS 45 CA using a custom pinna. Applying the model or target curve to headphones measured on a different test fixture will produce different scores that are likely not valid.
- Only use the Harman target as a guideline or reference when comparing relative differences among different headphones. While most people like it, it is not the last word on what sounds good. Program is a factor and recordings vary in spectrum. Try making broadband adjustments to the bass and treble using shelving filters (e.g. bass and treble tone controls). What sounds good to you is ultimately what matters.

15.7 NO CORRELATION BETWEEN PRICE AND FREQUENCY RESPONSE OR PREDICTED SOUND QUALITY

The current state of the art in headphone design was best captured by three large surveys of headphones available in the marketplace at the time: Breebaart (2017), Olive et al. (2018b) and Bems (2023a).

Breebaart (2017) published a large survey of 283 models of headphones with the provocative title “no correlation between headphone frequency response and retail price”. The survey included three headphone types: circumaural (AE), supra-aural (OE) and IE headphones. He calculated the correlation between headphone retail price and how far the frequency response deviated from two references: the mean response across all headphones, and an objective quality metric: the Harman target curve. High correlation values implied that spending more money would get you a better

sounding headphone. Using the Harman target curve as a reference the Pearson correlation coefficients for each type were: AE (r = 0.12), IE (r = 0.17) and OE (r = 0.02). The correlation values based on the mean response were very similar. The paper initially shocked the headphone community with the realization that price was no guarantee for better sound. It was a wake-up call to the consumers and the industry that they had no clear idea or consensus how to design headphones that sound good.

This author had already come to this conclusion in 2012, but Breebaart's study provided definitive proof. Inspired by his paper, we conducted a similar survey of 158 headphones that included the three types, both consumer and professional models (Olive et al., 2018b). The frequency responses of the headphones were used to predict their preference ratings using the two predictive models discussed in the previous section. The results are summarized in Figure 15.18. The left panel shows for each headphone type the average frequency response (black), the Harman target (thick gray) and the error curve (dotted). The dashed lines around the mean response represent ±1 standard deviation. Based on headphone type the average AE headphone comes closest to the Harman target but has 2–4 dB extra boost 60–300 Hz and 2–4 dB less energy between 3 and 8 kHz. OE headphones deviate from the Harman target in a similar manner, and IE headphones deviate even further over much of their bandwidth.

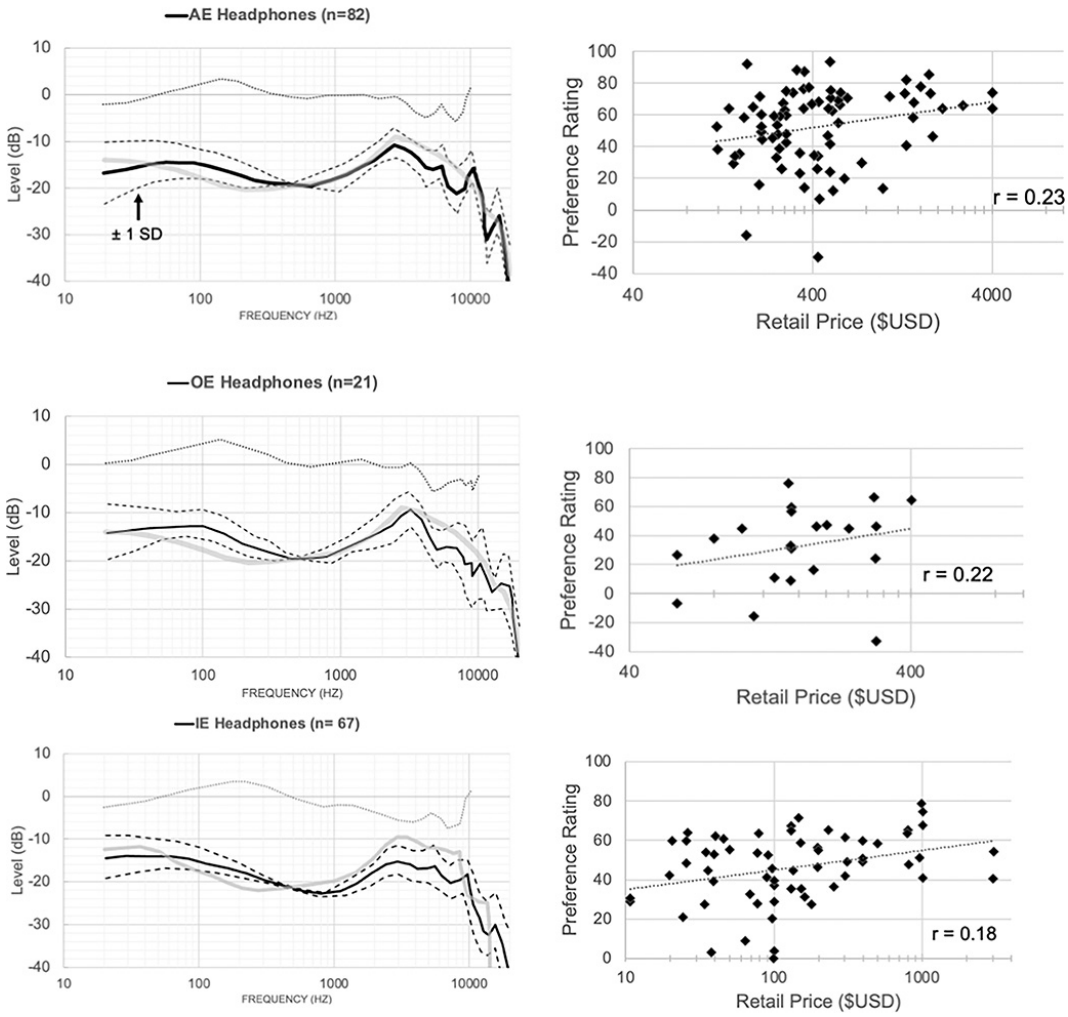


FIGURE 15.18 *Left panel: From top to bottom: the average frequency response of the tested samples for OE, AE and IE headphones (solid black curve) with ±1 standard deviation (dashed curves), and error curve (dotted curves) and Harman target curve (solid, thick gray). Right panel: The predicted preference rating for each model tested vs. its retail price with the best-fit regression line and correlations coefficient (Olive et al., 2018b).* ⏎

The right panel in Figure 15.18 plots the retail price of each model vs. its predicted preference rating. The dotted line is the best-fit regression line. The Pearson correlation coefficient *r* is an indicator of how well the data fits the line with a value of 1 indicating perfect correlation between price

and sound quality. Unfortunately, the correlation values are quite low ranging from 0.23 for AE headphones to 0.18 for IE headphones. *Conclusion: the retail price of the headphone is not a good indicator of its sound quality.*

Bems (2023a) performed a large headphone survey and found little consensus and consistency in their measured frequency response. A global weighted average frequency response using the top 162 models based on market share was calculated for each type. Three models from Apple accounted for 20% of the market share, meaning they are heavily weighted in this calculation. He plots the results for earbuds, IE, AE, OE and a Global average, which is a weighted average of all types (see Bems, 2023a, figure 15.5). The curves are referenced to a flat or neutral target curve which is not defined. All headphone types tend to have a smiley-face contour with a broad peak near 100 Hz and another one centered at 8 kHz, ranging in 2–6 dB from his reference "flat" target. The magnitude of the peaks varies with headphone type: AE and IE models are the most neutral, while OE are the worst.

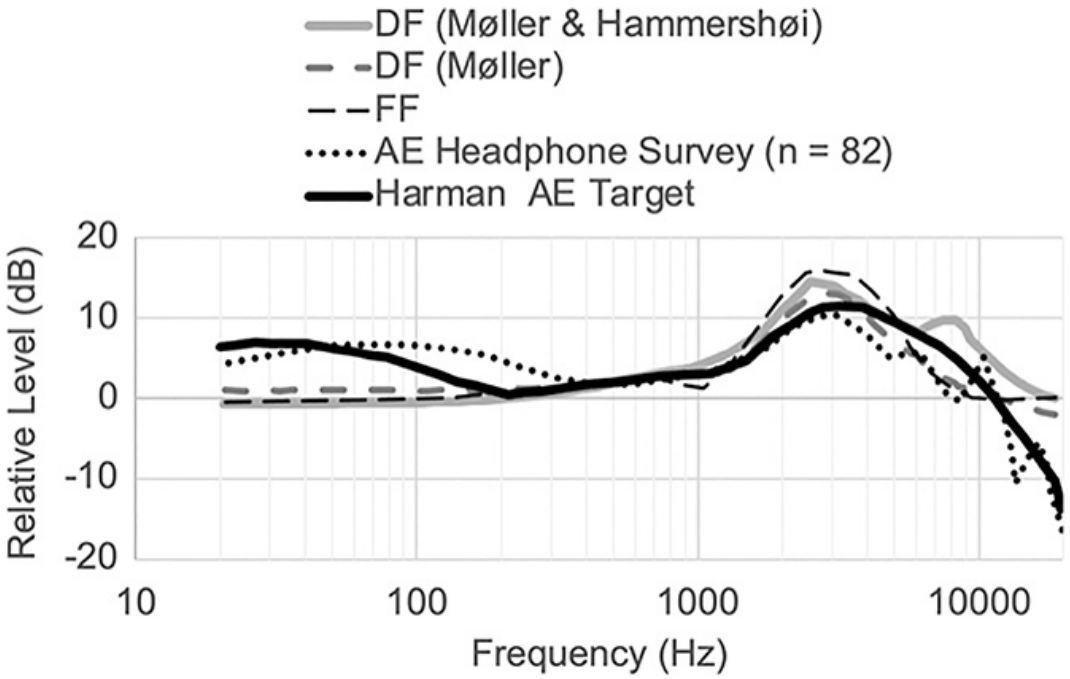


FIGURE 15.19 *Proposed target curves normalized at 500 Hz including two DF and FF targets, the Harman target and the*

average of 82 models reported in Olive et al. (2018b). ⏎

The lack of correlation between headphone price, frequency response and predicted sound quality reported in these three studies suggests the headphone industry has not yet caught up with recent advancements in scientific knowledge of what makes a headphone sound good. It would be interesting to repeat the study in 2024 to see if the quality and consistency has improved. What is clear is that the average headphone frequency response in 2018 was closer to the Harman target than the DF and FF target curves recommended by the current international standards.

Figure 15.19 shows the average response of the 82 AE headphones in Figure 15.18 compared to the Harman, DF, and FF target curves (DF and FF targets from Møller et al., 1995b). Although the DF and FF targets specify a flat response below 200 Hz, the average AE headphone and Harman targets have 5–6 dB more bass, which approximates the preferred in-room response of a full-range loudspeaker calibrated in a typical listening room. The average AE headphone has more energy between 200 and 400 Hz, which tends to produce lower preference ratings due to a perceived muddiness or tubbiness in vocals. The only consensus among headphone designers is which targets are *not* preferred.

15.8 HARMAN TARGET TESTED AGAINST OTHER PROPOSED TARGETS

Since the publication of the Harman research, other target curves have proposed by headphone researchers, reviewers and enthusiasts. Some have been validated with controlled blind listening tests that included the Harman target or some modified version of it. Publications of these test results are interesting to review since they provide further testing and validation of the Harman target from an independent third party.

15.8.1 Sonarworks

Sonarworks provides a software app intended for equalizing studio loudspeakers and headphones to their reference SoundID target curve. The curve is based on a flat near-field loudspeaker calibrated in a professional control room. Users can further fine tune the reference target to personal taste.

After performing A/B testing of 235 different target curves involving 147,000 individuals, they found that no one target satisfies most listeners and about 80% of listeners would benefit from some form of headphone personalization (Bems, 2023b). However, 40% of the listeners can be satisfied with adjustments in the bass/mid/treble of their target that fall within ±3 dB.

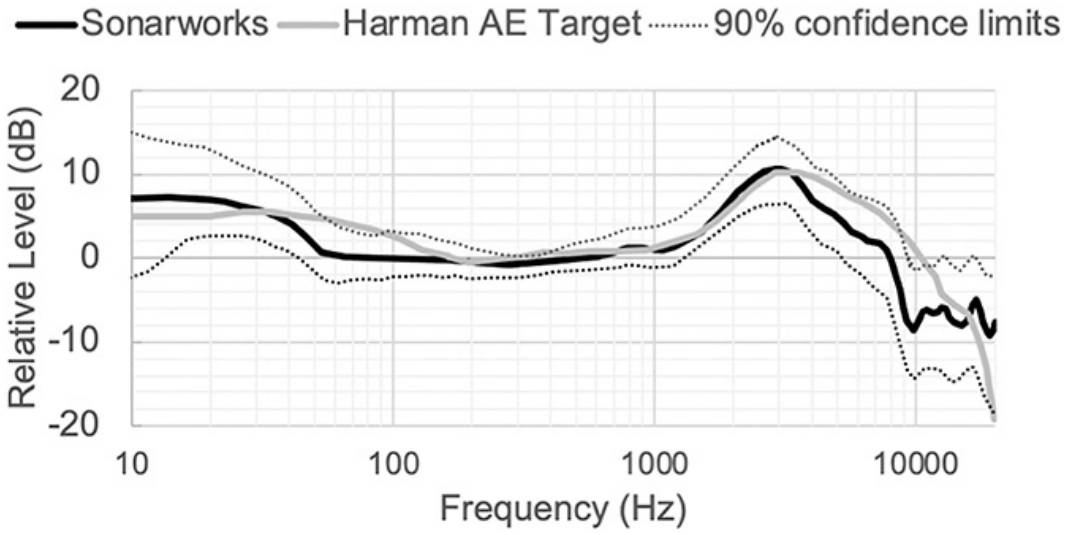


FIGURE 15.20 *The estimated Sonarworks Sound ID target curve (black) relative to the Harman AE target curve (light gray). The estimation is based on the mean frequency response of 122 different models of headphones equalized to the target using Sonarworks software and measured on a GRAS KEMAR 45BC. The dotted curves are the 90% confidence intervals. (Special thanks to Konstantin Davy, the author of this work.)*

Both the SoundID target curve and the headphone test fixture used to measure it are proprietary. However, the target curve was estimated on a

GRAS 45BC-10 KEMAR by measuring 122 headphones equalized to the SoundID target via their software app and calculating the average response. Figure 15.20 shows the estimated SoundID target compared to the Harman AE/OE target. The dotted lines represent the 90% confidence intervals around the mean response. The Harman target falls within the confidence limits and there is good agreement between 100 Hz to 2 kHz. The low-frequency shelf of the SoundID target begins at a lower frequency (60 Hz) and has 2 dB more gain. It also has more gain between 2 and 3 kHz and significantly less gain between 3 and 10 kHz. In terms of spectral balance, the SoundID target should sound less warm and full in the upper bass and less bright in the upper mids and treble compared to the Harman target. However, it is important to remember that this is only an estimation of the SoundID target. Even so, the wide confidence limits suggest that consumers who equalize their headphones to the SoundID target will get different results depending on which headphone model they own and how well its frequency response matches the sample(s) Sonarworks used to calculate the EQ correction.

15.8.2 Knowles IE Headphone Target Curve

Miller and Downey (2023) conducted listening experiments to study whether the Harman IE target curve (Olive et al., 2019) could be improved beyond 8 kHz. Since the Harman research used a Type 3.3 ear simulator, they argued the accuracy and reproducibility of the measurements beyond 10 kHz may have been a limiting factor. Moreover, the preferred treble levels beyond 8 kHz may depend on the age and hearing of the listeners, something we suspected but had no audiometry data to provide definitive proof.

High-frequency audiometry tests revealed the listeners had the expected increased hearing thresholds as a function of age (see Figure 15.21(a)). Using a method of adjustment and several tracks of music they found the preferred target for the youngest group of listeners with normal hearing was 12 dB higher at 16 kHz than the Harman target. The preferred high-frequency adjustments to the Harman target are shown in Figure 15.21(b) as

function of age. For every additional 10 years of age over 30, the preferred treble level increases 3 dB.

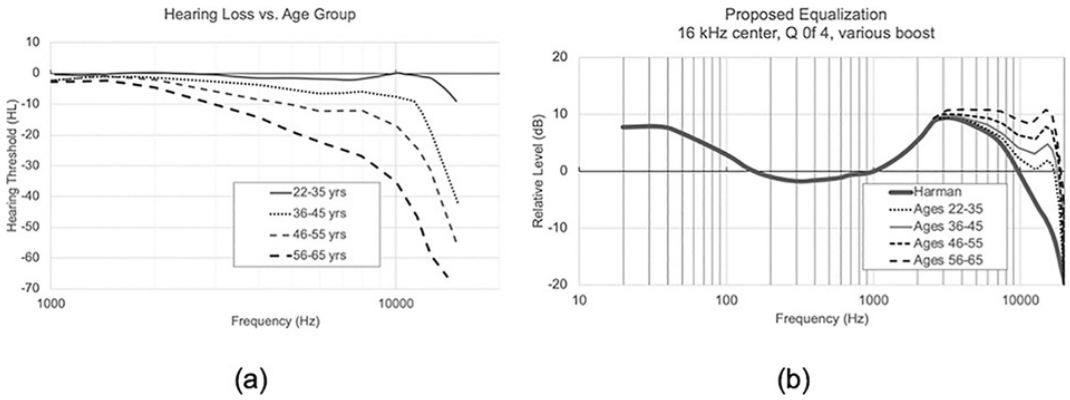


FIGURE 15.21 *(a) The change in hearing thresholds with age, relative to the group of 10- to 21-year-old listeners. (b) The preferred headphone adjustments to the Harman IE target curve based on age (Olive et al., 2019). Figures 9 and 12 from Miller and Downey (2023) redrawn.* ⏎

In Sections 15.6.4 and 15.6.5 it was noted that older listeners tended to prefer headphones with less bass and more treble and that this could perhaps be compensation for age-related hearing loss. The Knowles research provides direct evidence that increased hearing loss leads to an increase in the preferred treble level.

15.8.3 SoundGuys

SoundGuys (www.soundguys.com) is a headphone review site based in Vancouver, Canada, that provides headphone measurements made on a B&K Type 5128-B. Their target curve is based on the average frequency response of several popular reference headphones from each category (AE/OE/IE) that they feel are among the best sounding headphones. They have not conducted formal blind listening tests to support these claims.

They also provide a “legacy” Studio target curve aimed at professional audio recording mixers and producers. This target is based on an average measurement of IE and head-band types from Sennheiser, Beyerdynamic

and other professional headphones. The Studio target is closer to the DF target and is flat in the bass. It is a common notion that professionals prefer to mix on loudspeakers and headphones that have less bass than what consumers prefer.

Figure 15.22 plots both the consumer (labeled as SoundGuys) and SoundGuys Studio target curves. Also shown is the Harman (5128 modified), their estimation of the Harman AE/OE target translated to a B&K Type 5128. The consumer target curve is very similar to the Harman target up to about 2 kHz and drops 2–3 dB below the Harman target between 3 and 6 kHz. As mentioned earlier, SoundGuys has not published results of any formal listening tests on their headphone target. However, Olive (2024) included their target in controlled listening tests and it was equally preferred to the Harman target (see Section 15.6.9). SenseLab (Ravizza et al., 2023) also included their target in tests, which are reported in Section 15.8.5.

SoundGuys have recently supplemented measurement reports with perceptually derived scores from Head Acoustics Multidimensional Audio Quality Score (MDAQS). Simply explained it works as follows: binaural recordings are made of headphones on a Type 4.3 HATS reproducing both technical and music signals. The recordings are then analyzed using a perceptual model that predicts sound quality scores on a 5-point scale for three attributes: timbre, immersion and distortion. An overall quality score is calculated based on a weighted average of the three attributes. More details are available at Wykes (2024b).

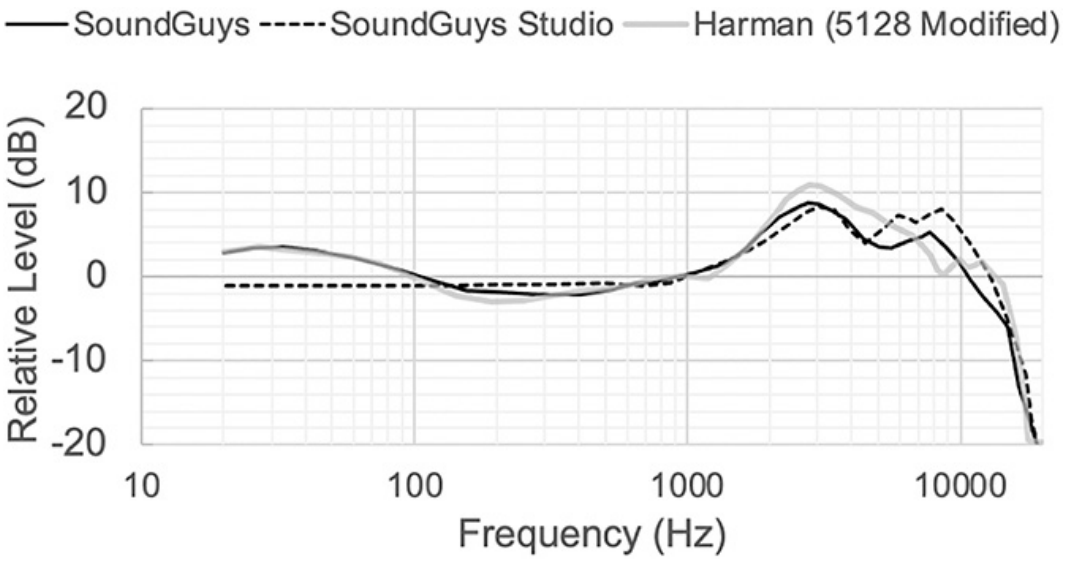


FIGURE 15.22 *The SoundGuys.com target curves for consumer (black) and professional (dotted) headphones compared to the Harman 2015 AE target—all defined on the B&K Type 5128.*

The author has no first-hand experience using MDAQS. However, the appeal of using perceptual-based models to predict sound quality is undeniable given its potential to save considerable time and money running listening tests.

15.8.4 Rtings

Rtings (www.rtings.com) is based in Montreal, Canada, and publishes objective-based measurements and test reports on headphones, sound bars and other consumer products. Their headphone target is Harman-like but modified for a Head Acoustics HMS II.3 (Vafaei, 2018). Recently the fixture was replaced with a B&K 5128-B, and a study was conducted to define a new target curve specifically for the fixture. They conducted 350 controlled listening tests where five different target curves were rated based on preference. The listening panel included 35 employees—12 of whom were audio professionals (Lelièvre, 2024). A virtual headphone method was used where the replicator headphone (Sennheiser model 650) was equalized to each of the targets. Two short music loops were used, and level

adjustments were made to remove loudness differences among the five targets.

Figure 15.23(a) plots the five targets they tested:

1. *DF Tilted*: The B&K 5128 DF response with a −1 dB per octave tilt.
2. *SenseLab*: A visual approximation of the average preferred curve in the Ravizza et al. study (2023).
3. *Rtings SPK*: A target based on the Rtings in-room target for soundbars.
4. *DF Tilt + Harman Bass*: The B&K 5128 DF response with a −1 dB per octave downward tilt at 200 Hz with a bass shelf that approximates the Harman AE/OE target.
5. *Low Bass*: A low anchor intended to calibrate the scale and help listeners use the scale more consistently. A −4 dB bass shelf below 300 Hz and a −5 dB dip centered at 3 kHz.

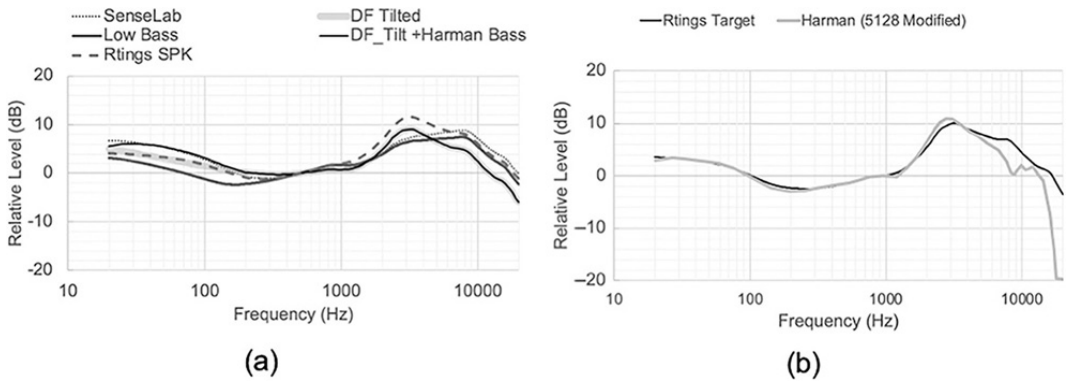


FIGURE 15.23 *(a) The five target curves evaluated by Rtings in controlled listening tests. (b) The final target compared to the Harman AE 2015 (modified to 5128). All curves are measured on the B&K Type 5128-B (Lelièvre, 2024).* ⏎

The first four target curves are all variations of what most listeners consider neutral. The *Low Bass* anchor is a low sound quality target intended to anchor the scale. As expected, the results revealed the first four targets were statistically tied, while the *Low Bass* target was rated

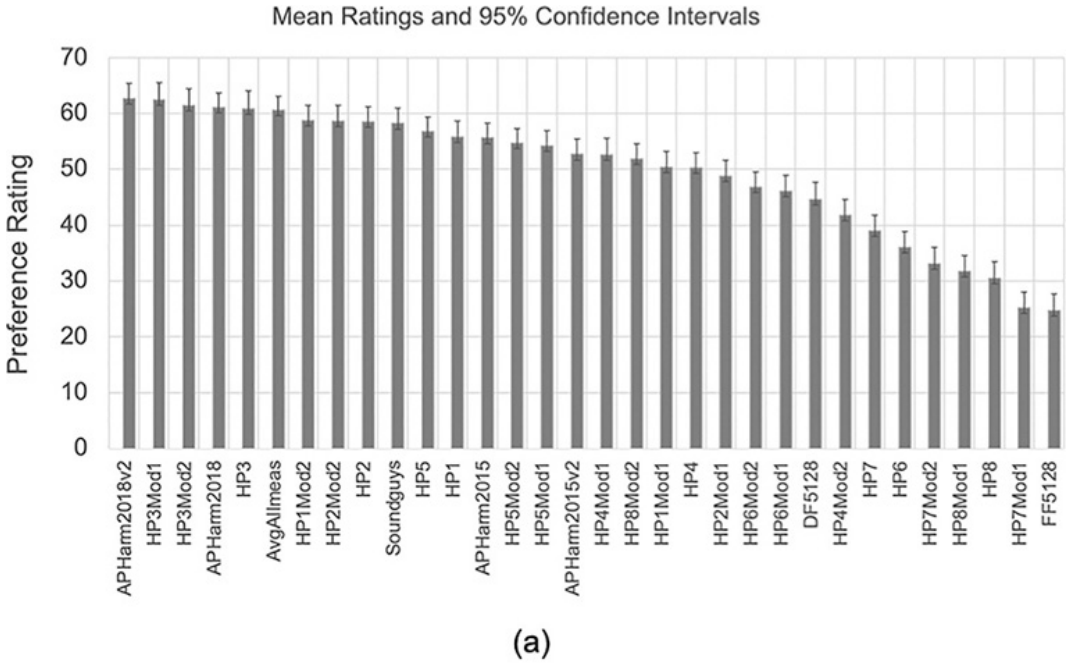
significantly lower and not preferred. The conclusion was that although the top four target curves sounded different from each other, they all sounded acceptable to the average listener. They noted that women tended to rate the *Low Bass* target higher than men. In two studies discussed in 15.6.4 and 15.6.5, we noted that women tended to prefer headphones with less bass than men.

Figure 15.23(b) shows the final choice of their reference target curve. It is based on the 1/6-octave-smoothed B&K 5128 DF Target with a −0.6 dB per octave downward tilt and a 4.3 dB bass shelf boost 105 Hz (Q = 0.707). It approximates the Harman (5128 modified) target and the PIR curve of a well-designed loudspeaker like the Revel F208 (see Figure 15.7(d)). Moving forward, Rtings intends to incorporate a feature on their website where users can plot measurements of headphones available in their database and include their favorite target curve as a reference. From this they can find which headphones are the closet match to their preferred sound profile.

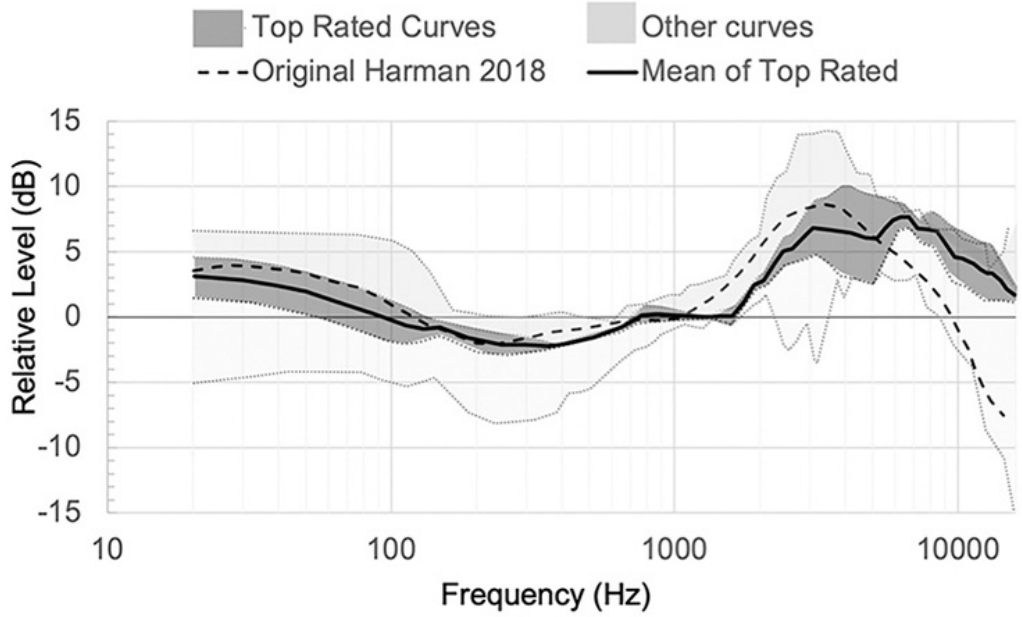
15.8.5 SenseLab Study

SenseLab and the University of Aizu in Japan conducted a joint study in 2022 where 56 untrained listeners across both countries evaluated 32 different headphone target curves using many tracks of music (Ravizza et al., 2023). A virtual headphone method was used where different equalizations were applied to a replicator headphone (Beyerdynamic DT770 Pro).

The headphones measurements and equalizations were made on a B&K Type 5128. The test included virtual versions of eight popular closed-back headphones with two additional versions of each where the bass or treble was adjusted. The average frequency response of the eight headphones was included for a total of 25 equalizations. The remaining seven virtual headphones included well-known industry target curves including the B&K 5128 FF and DF targets, SoundGuys consumer target, two approximations of the Harman AE Targets (2015 and 2018) provided by a third-party, and their modified versions.



(a)



(b)

FIGURE 15.24 *(a) The mean preference rating and 95% confidence intervals for 32 headphone curves averaged across all listeners and programs from study by Ravizza et al. (2023). (b) The frequency response of two groups of headphones based on preference ratings. The top five rated headphones fall within the dark shaded area and the remaining 27 headphones in the light*

shaded area. The solid line is the mean of the top five headphones and the dotted line is the Harman 2018 target measured on the GRAS45CA. All other curves shown here were measured on the B&K 5128. ⏎

Figure 15.24(a) summarizes the mean preference ratings and 95% confidence intervals for the 32 curves tested. The following conclusions were drawn from the study:

- There were no significant effects in preference related to cultural differences (Danish vs. Japanese).
- The music tracks influenced which target was preferred.
- While one of Harman target curves was rated highest overall, the differences in mean ratings among the 11 top rated headphone targets were not statistically significant ($p \leq 0.05$).
- The lowest rated headphone targets were the FF and DF target curves, both industry recommended standards.
- Segmentation analysis revealed no clear clusters based on differences in preferred sound profile.

Figure 15.24(b) illustrates the measurements of the headphones according to how much they were preferred. The five top-rated headphones fall within the narrow dark shaded area with the mean response of the group indicated by the black curve. The remaining 27 curves fall within the light shaded area. The dotted curve is the original Harman 2018 target provided by a third party. It doesn't appear to have been translated from a Type 3.3 to a Type 4.3 fixture used in this study. What is remarkable is that when the curves are normalized at 500 Hz, the preference tolerance are quite narrow between 300 Hz to 2 kHz, the most important range for speech and music. The bass and treble regions are still quite narrow, ±2 dB below 125 Hz, and above 8 kHz with the broadest tolerances ±3.5 dB falling between 3 and 5 kHz.

Overall, this study provides an important contribution to the scientific literature related to the perception and measurement of headphone sound quality. It also provides further independent confirmation that the Harman AE/OE target is well accepted by both Danish and Japanese listeners.

15.8.6 Headphones.com: Preference Bounds and DF Calibration

Headphones.com (www.headphones.com) is yet another Canadian headphone review site based in Vancouver that provides headphones measurements on both Type 3.3 and Type 4.3 (B&K Type 5128-B) test fixtures. While their subjective reviews of headphones are not based on controlled listening tests, their approach towards measurements and their interpretation of them is firmly rooted in science

Headphones.com doesn't recommend a specific target curve but rather suggests a preference bounds that the headphone frequency response should fall within to statistically satisfy most tastes (see Silver, 2024b, figure 8). The idea was inspired by the segmentation research (see Section 15.6.7) and recognizes that a single target curve cannot satisfy all tastes.

Headphones.com advocates DF-calibrated measurements whereby the DF response of the test fixture is subtracted from the raw measurement (see Silver, 2024b, figure 12). This removes the transfer function of the test fixture from the measurement and in theory makes it easier to compare measurements of headphones made on different test fixtures. Because the DF response of the test fixture is determined using a loudspeaker, DF-calibration will not account for and remove the fixture's idiosyncratic acoustic source impedance interactions with headphones from the measurement (see Figures 15.19 and 15.20, Silver, 2024b). Measurements of IE headphones bypass the acoustic effects of the pinna and head so applying the DF-calibration to them will produce somewhat different results.

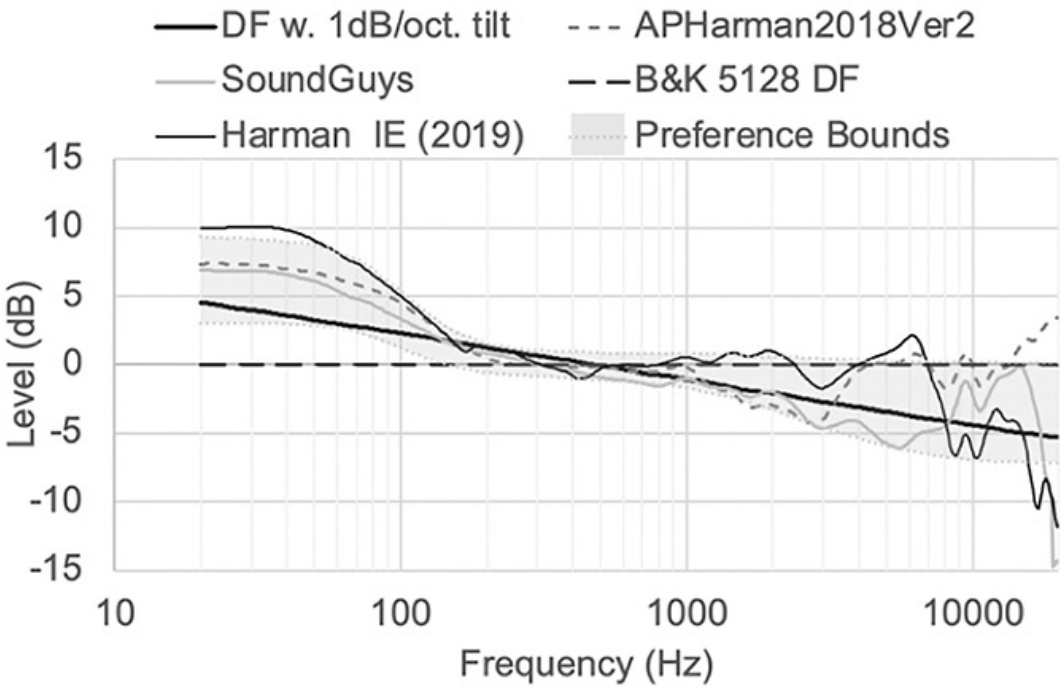


FIGURE 15.25 *An example of four headphone target curves where the raw response measured on the B&K Type 5128 have been compensated by its DF response with Preference Bounds suggested by Silver (2024b).* ⏎

Figure 15.25 is an example of several headphone targets plotted with preference bounds after DF calibration for the B&K 5128 has been applied. These target curves have all produced high preference ratings in controlled listening tests reported in the previous sections and fall within the preference bounds. The exception here is the DF target (dashed line at 0 dB). It does not receive high preference ratings in listening tests and below 100 Hz falls outside the bounds indicating insufficient bass. Above 100 Hz it sits just on the edge of the upper bounds. The DF target with a −1 dB/octave downward tilt lies in the middle of the bounds and is a good match to the PIR of a well-designed loudspeaker like the Revel F208. Loudspeakers with measurements like this also receive high preference ratings in controlled listening tests.

Finally, it is a matter of debate whether DF-calibrated measurements that produce downward tilted curves are easier to interpret than error curves (the

difference in responses between a target curve and the headphone; see Figure 15.13) that aim for a flat curve. Most audio people instinctively understand that flat equals better sound. Adding a preference bounds to the error curve indicate the extent to which it satisfies the tastes of other listener segments identified in Section 15.6.7.

15.9 THE THREE FACES OF HARMAN TARGETS

Between 2012 and 2019, Harman research produced three AE/OE and three IE headphone targets. The large number of Harman targets has been a source of confusion and frustration for some headphone enthusiasts who complain that if science is the truth, there should only be one answer. Truth in science is based on the best evidence available at the time, and with further investigation comes new evidence. And so it goes with the science of headphone sound quality. The truth is a moving target—or in our case, three targets.

Over the course of 7 years, Harman conducted multiple listening experiments that produced new evidence and incremental improvements in the preferred the target curve. Figure 15.26(a) plots the three different target curves for AE/OE headphones published between 2013 and 2017, all measured on the GRAS 45CA MOD and normalized to 500 Hz. The 2013 AE/OE headphone target was a result from an experiment where listeners adjusted the bass and treble of the Harman baseline target: the flat in-room response of the loudspeaker (Olive et al., 2013b). The 2015 target curve was born from a much larger study where 258 listeners from four different countries adjusted the bass and treble of the baseline target based on preference (Olive et al., 2015). This target is more generalized to a wider listening population (Section 15.6.5). After adjustments the 2015 target saw a 2 dB increase in the bass and treble relative to the 2013 target. The bass and treble remained relatively unchanged in the 2018 target but it saw a 3 dB reduction in a frequency band centered at 3 kHz, the sensitive region where the ear canal resonance occurs. This modification was in response to feedback from unpublished listening tests results. The 2018 target was

tested against 31 headphones using 130 listeners and on average received the highest preference ratings (see Section 15.6.6).

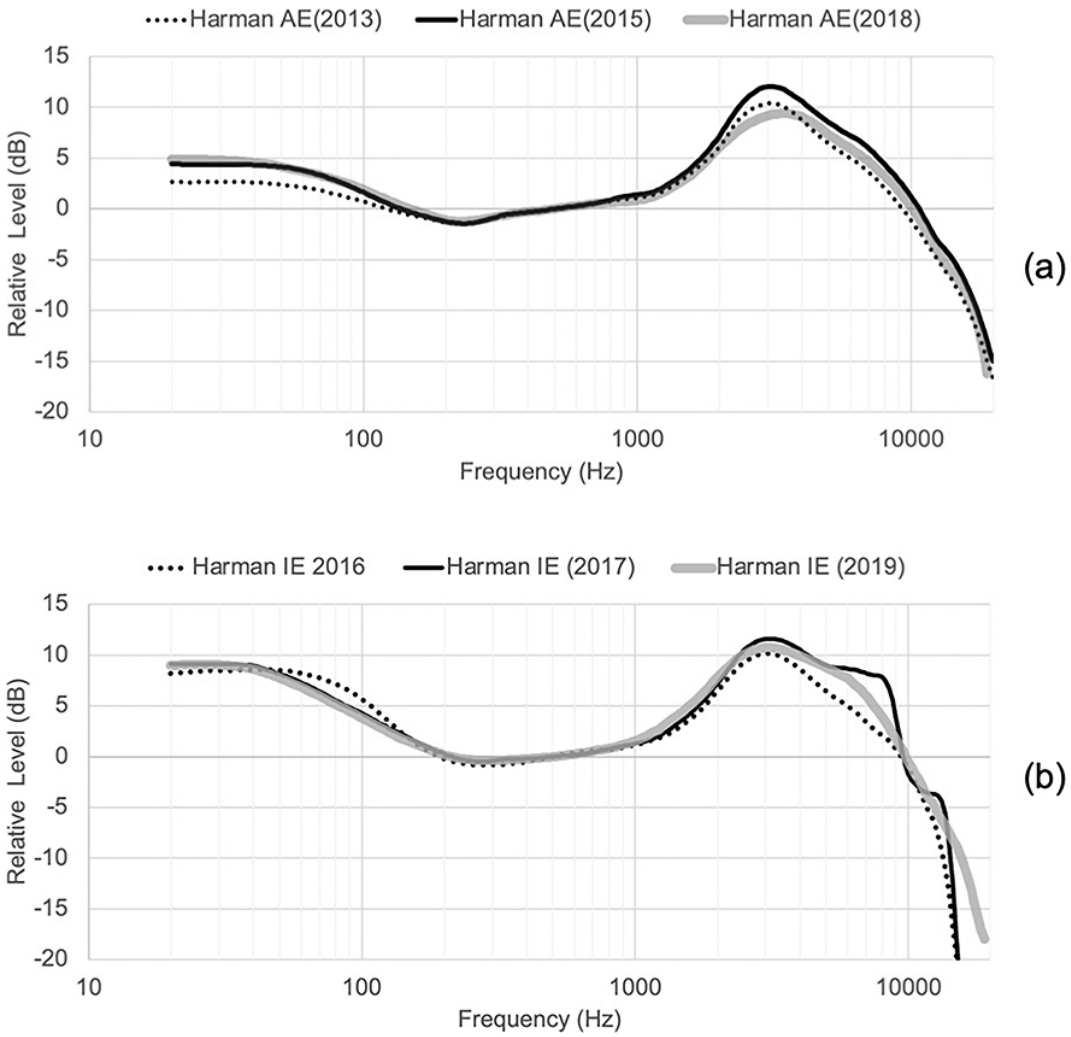


FIGURE 15.26 *(a) The evolution of the AE/OE Harman target curves and (b) IE target curves from 2013 to 2019 based on experiments and listener feedback. All targets shown are based on headphones measured on a Type 3.3 GRAS 45CA MOD.* ⏎

Figure 15.26(b) shows the 2016 IE target curve which began as the 2015 AE target curve with 4 dB of additional bass to make it sound similar in balance. Listening experiments (Olive et al., 2016a) focused on the preferred bass shelf filter which was applied to the 2017 IE target curve. This target tested against 31 other IE headphones that was preferred overall

(Olive et al., 2017a). The final 2019 IE target removed the bump at 8 kHz, after we determined it was perceived as sounding more neutral.

In summary, the different faces of the Harman target curves can be traced to incremental improvements based on new scientific evidence and feedback from listening tests. The measured differences between versions are in absolute terms quite small varying no more than 2 dB in the bass and 1–2 dB in the treble. Differences of these magnitude are probably within manufacturing tolerances and the preference bounds shown in Figure 15.25: either one of the curves would likely satisfy most listeners. In 2019, we came to the realization that nuisance variables like program material, individual variations in frequency response from leakage/ear coupling and listener segmentation in preferred sound posed much larger challenges to delivering a consistent level of sound quality in headphones than any incremental improvements realized in further target research. Our efforts were better spent finding solutions to these problems.

15.10 HEADPHONES FOR SPATIAL AUDIO

15.10.1 Growth in Immersive Audio

So far, this chapter has only considered headphone reproduction for stereo content. However, spatial audio content intended for binaural reproduction through headphones is available for games, movies and music and augmented and virtual reality applications. Headphones are the most common device through which spatial audio is experienced with 39% (Bull and Duval, 2023) of the content from gaming, AR/VR applications, music, film, television and podcasts rendered through headphones. In automobiles, binaural audio reproduction of content through near-field loudspeakers mounted in the headrest is expected to grow as manufacturers are looking for ways to reduce the weight, cost and complexity of audio systems, while providing individual sound zones that allow occupants to enjoy different content at the same time. The key enablers for wider adoption are (1) the content has become easily accessible from streaming services like Netflix,

Apple Music, Amazon HD and Tidal; and (2) the content is scalable to the playback system and can be rendered to headphones or whatever number of speakers are available.

15.10.2 Is There a Different Headphone Target for Spatial Audio?

An obvious question is whether the headphone target for stereo is appropriate for reproduction of spatial or binaural content. Researchers at Reality Labs, Meta (Engel et al., 2022) studied this question conducting separate tests where 20 listeners compared stereo and binaural content through headphones equalized to different targets. The targets included the Harman target (Olive et al., 2013a), as well as bass and treble modifications made to it, and a flat target curve which they hypothesized would be preferred for binaural rendering. The targets were evaluated using both generic and individualized headphone equalization measured at the blocked entrance of listeners' ear canals. Each listener's binaural room impulse response (BRIR) was captured from two different pairs of high-quality professional monitors in two rooms: an apartment and small control room. The four stereo tracks were then convolved with the individual BRIRs to create the spatial audio content, and the unprocessed tracks were used for the stereo content.

The flat target was preferred for binaural content and the Harman target, for stereo content. Personalized headphone equalizations produced the highest preference ratings as did binaural content, although the trends remained the same. The effect of content on preference was smaller for the non-individualized calibration.

The flat target for binaural audio agrees at least two other studies by Møller et al. (1996) and Pralong and Carlile (1996). However, in recent discussions with two spatial audio researchers they preferred a prefer binaural target that had a bass shelf boost like a loudspeaker in a room. The bass boost may also compensate for the missing 3 dB due to missing haptic information when moving from loudspeaker to headphone listening (see Section 15.1.4).

In the Engel paper, the loudspeaker was equalized to have a flat, steady-state in-room response, which would boost the direct sound above 2 to 3 kHz and make it sound bright. The Harman target would add additional gain in this frequency region, and this could explain why it was less preferred to a flat target for spatial audio content.

15.11 TOPICS FOR FUTURE RESEARCH

Most of the research presented so far has focused on the perception and measurement of headphone sound quality. Frequency response has been identified as the single most important parameter related to headphone sound quality, and there is some consensus that a target curve defined with some tolerances around it will satisfy most tastes. Here is a list of suggested research topics related to the perception and measurement of headphone that could use more attention:

- *Nonlinear distortion*: At what level does it become audible and affect sound quality and are there measurements that can reliably predict this?
- *Spatial attributes*: Are there objective-based measurements that can predict the perceived spatial attributes of audio reproduced through headphones including localization, externalization and envelopment?
- *Individual ear canal acoustics*: The varied geometry and acoustic impedance of human ear canals affect the magnitude and phase response at the ear drum. How much does this effect perceived timbre and spatial quality of headphones? Can we model, predict and compensate for this?
- *Hearing assistance and protection*: Products like the Apple AirPods Pro 2 have built-in hearing compensation for people with mild-to-moderate hearing loss. In hear-through mode they can limit exposure to external sounds to about 85 dB (Butterworth, 2024) up to a point. These features will likely become more important to consumers, and

more research will likely focus on how to optimize and measure their effectiveness.

- *Perceptual models*: Perceptual models that predict sound quality including its attributes (timbre, spatial, dynamic, distortion) and account for hearing loss, program signals and background noise and other variables will lead to more robust and generalized predictions of sound quality. Such models can help make the design and testing of headphones more efficient and optimized and are key to better personalization of headphone sound quality.
- *Personalization*: Are there more efficient (automated) and effective ways to personalize a headphone to accommodate individual differences in taste, hearing loss, and automatic adjustment based on listening context: quiet vs. noisy listening, activity (e.g. home, gym, outdoor, cycling), private vs. social mode?

These are interesting topics, and hopefully some enterprising headphone researchers will find the motivation and support to tackle them.

15.12 CONCLUSIONS

Headphones have become the primary devices through which most audio content is heard today, yet their sound quality has only recently received the attention it deserves. Experimental evidence from several independent studies indicate that listeners prefer headphones with a frequency response that approximates the predicted in-room response (PIR) of a neutral loudspeaker calibrated in a semi-reflective listening room. Since music recordings are optimized for loudspeaker reproduction in rooms like this, it makes sense that the best-sounding headphones approximate these conditions.

The Harman target curve is one example that satisfies most (64%) listeners, and the tastes of two smaller listener segments can be satisfied with simple adjustments to the bass and treble. Certain demographic factors

(age, gender, listening experience) tend to be associated with membership in the three segments. Older, more experienced listeners tend to prefer less bass than younger, less experienced listeners, Males tend to prefer more bass and treble than females. Hearing loss seems to be associated with a preference for more treble and less bass.

Headphone companies have an opportunity to personalize headphones in order to accommodate these different tastes. Headphone reviewers are beginning to recognize that a single target cannot satisfy everyone's taste. Some are plotting the headphone's frequency response within a preference bounds or mask that defines what is considered acceptable by most listeners using evidence from our research. Other sites allow the user to specify their preferred target curve and search the database of headphone measurements to find the best match.

The optimal design and testing of headphones are complicated by variations in the size and shapes of human heads and ears and how they acoustically couple to the headphones. These variations can produce large differences in bass (leakage effects) and treble that are often not accurately captured by the headphone test fixture. The bottom line is that the measurement of the headphone may not accurately predict what the listener hears. The best headphones designs minimize these variations and deliver the same sounds consistently across a range of listeners.

Much has been learned over the past decade about the perception and measurement of headphone sound quality. One lesson learned is that DF target recommended in the current international standards is consistently rated lower in sound quality than alternative targets. It might be time for the headphone standards to catch up with modern science.

REFERENCES

ANSI/ASA S3.36 (2018). "Manikin for Simulated In-Situ Airborne Acoustic Measurements". ⏎

Apple (2024). "AirPods Pro 2". www.apple.com/airpods-pro/hearing-health/# ⏎

Austrian Academy of Science (2024). www.oeaw.ac.at/en/ari/das-institut/software/hrtf-database ⏎

Begault, D., Wenzel, E. M. and Anderson, M. R. (2001). "Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source", *J. Audio Eng. Soc*., vol. 49, no. 10, pp. 904–916. ⏎

Bems, H. (2023b). "SoundID SR for Headphones: A New Headphone Tuning Standard". www.sonarworks.com/blog/research/white-paper ⏎

Bems, H. (2023a). "Headphone Sound Report". www.sonarworks.com/blog/research/headphone-sound-report-2023 ⏎

Blauert, J. (1983). *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA. ⏎

Breebaart, J. (2017). "No Correlation between Headphone Frequency Response and Retail Price", *J. Acoustical Soc. Am*., vol. 141, no. 6, EL526–EL530. https://doi.org/10.1121/1.4984044 ⏎

Briolle, F. and Voinier, T. (1992). "Transfer Function and Subjective Quality of Headphones: Part 2, Subjective Quality Evaluations", *11th Audio Eng. Soc. International Conference*. ⏎

Bull, C. and Duvall. (2023). "Audio Tech Lifestyles Consumer Survey: USA, UK, Germany, Japan, China", *FutureSource*. ⏎

Butterworth, B. (2024). "New Ways to Measure Noise Cancelling, Transparency, and Other Modes in Headphones and Earphones", *Audioxpress*. https://audioxpress.com/article/new-ways-to-measure-noise-cancelling-transparency-and-other-modes-in-headphones-and-earphones ⏎

Darkner, S., Sommer, S., Schumacher, A., Ingerslev, H., Baandrup, A., Thomsen, C. and Jonsson, S. (2018). "An Average of the Human Ear Canal". https://doi.org/10.48550/arXiv.1811.03848 ⏎

Dobrucki, A., Kin, M. and Kruk, B. (2013). "Preliminary Study on the Influence of Headphones for Listening Music on Hearing Loss of Young People", *Arch. Acoust*., vol. 38, no. 3. http://doi.org/10.2478/aoa-2013-0045 ⏎

Engel, I., Alon, D., Scheumann, K. and Mehra, R. (2022). "On the Differences in Preferred Headphone Response for Spatial and Stereo Content", *J. Audio Eng. Soc*., vol. 70, no. 4, pp. 271–283. ⏎

Fasanya, B. and Strong, J. (2019). "Younger Generation Safety: Hearing Loss and Academic Performance Degradation among College Student Headphone Users", in Arezes, P., editor, *Advances in Safety Management and Human Factors: AHFE 2018, AISC 791*, pp. 522–531, Springer International Publishing AG, Part of Springer Nature. https://doi.org/10.1007/978-3-319-94589-7_51 ⏎

Fleischmann, A., Silzle, F. and Plogsties, J. (2012). "Identification and Evaluation of Target Curves for Headphones", *133rd Audio Eng. Soc. Convention*, Paper 8740. ⏎

Hammershøi, D., Reuter, K. Ordoñez, R., Mario de Santis, E. (2008). "Review of Literature om Hearing Damage by Personal Stereo", *Joint Baltic-Nordic Acoustics Meeting 2008*. ⏎

Hirvonen, T., Vallgamaa, M., Backman, J. and Karjalainen, M. (2003). "Listening Test Methodology For Headphone Evaluation", *114th Convention Audio Eng. Soc*., Paper 5736. ⏎

Hodgetts, W. E., Szarko, R. and Rieger, J. (2009). "What Is the Influence of Background Noise and Exercise on the Listening Levels of iPod Users?", *Int. J. Audiol*., vol. 48, no. 12, pp. 825 ± 32. https://doi.org/10.3109/14992020903082104 ⏎

IEC 60268-7 (2010). *Sound System Equipment—Part 7: Headphones and Earphones*, Geneva, Switzerland. ⏎

IEC 60318-4:2010 (2010). "Electroacoustics—Simulators of Human Head and Ear—Part 4: Occluded-ear Simulator for the Measurement of Earphones Coupled to the Ear by Means of Ear Inserts". https://webstore.iec.ch/en/publication/1445.

IEC 60318-7 (2022). "Electroacoustics—Simulators of Human Head and Ear—Part 7: Head and Torso Simulator for the Measurement of Sound Sources Close to the Ear". https://webstore.iec.ch/en/publication/67522 ⏎

IEC 60628-24 (2023). “Sound System Equipment—Part 24: Headphones and Earphones—Active Acoustic Noise Cancelling Characteristics”. https://webstore.ansi.org/standards/iec/iec6026824ed2023 ⏎

IEC 60711:1981 (1981). “Occluded-ear Simulator for the Measurement of Earphones Coupled to the Ear-by-ear Inserts”. https://webstore.iec.ch/en/publication/17389 ⏎

ITU-R BS. 708 (1990). “Determination of the Electroacoustical Properties of Studio Monitor Headphones”, *Report ITU BS.708, International Telecommunication Union*, Geneva, Switzerland. ⏎

ITU-R BS. 1116-3 (2015). “Methods for the Subjective Assessment of Small Impairments in Audio Systems”, *Report ITU BS.1116, International Telecommunication Union*, Geneva, Switzerland. ⏎

ITU-T P. 57 (2021). “P Series: Telephone Transmission Quality, Telephone Installations, Local Line Networks P. 50-P.59: Objective Measuring Apparatus: Artificial Ears”. www.itu.int/ITU-T/recommendations/rec.aspx?rec=14599. ⏎

ITU-T P. 58 (2023). “P.58: Head and Torso Simulator for Telephonometry”. www.itu.int/rec/T-REC-P.58/en

JBL (2024). “Safe and Fun Sound for Little Ears: JBL’s New Junior Headphones”. https://news.jbl.com/en-CEU/242001-safe-and-fun-sound-for-little-ears-jbl-s-new-junior-headphones’ ⏎

Johansen, B., Rasmussen, P. and. Pedersen, M. H. (2024). “Unveiling the Impact of Ear Canal Geometry for In-Ear Headphone Testing”, *Audioxpress*, June 26. https://audioxpress.com/article/unveiling-the-impact-of-ear-canal-geometry-for-in-ear-headphone-testing ⏎

Jønsson, S., Schuhmacher, A. and Ingerslev, H. (2018). “Wideband Impedance Measurement in the Human Ear Canal; In Vivo Study on 32 Subjects”, *Brüel & Kjær Sound & Vibration*. http://arxiv.org/pdf/1811.03389 ⏎

Kallinen, K. and Ravaja, N. (2007). “Comparing Speakers Versus Headphones in Listening to News from a Computer—Individual Differences and Psychophysical Responses”, *Comput. Hum. Behav.*, vol. 23, no. 1, pp. 303–317. ⏎

King, R., Theriault, M. and Massenburg, G. (2023). “A Practical Approach to the Use of Center Channel in Immersive Music Production”, *155th Convention Audio Eng. Soc*., Paper 168. ⏎

Kulkarni, A. and Colburn, H. S. (2000). “Variability in the Characterization of the Headphone Transfer-function”, *J. Acoust. Soc. Am*., vol. 107, no. 2, pp. 1071–1074. http://10.1121/1.428571. ⏎

Lelièvre, P. (2024). “One Curve Doesn’t Fit All: 350 Headphone Listening Tests Show That Many Sound Profiles Can Be Preferred”. www.rtings.com/research/frequency-response-1-8 ⏎

Lieberman, A., Schroeder, J. and Amir, O. (2022). “A Voice Inside My Head: The Psychological and Behavioral Consequences of Auditory Technologies”, *Organ. Beh. Hum. Decis. Process*, vol. 170. ⏎

Listen HRTF Database (2002). http://recherche.ircam.fr/equipes/salles/listen/ ⏎

Lorho, G. (2009). “Subjective Evaluation of Headphone Target Frequency Responses”, *126th Convention Audio Eng. Soc*., Paper 7770. ⏎

Mansilla, J. (2020). “Headphone Measurements: The New Standard, Part 1”, *Head-Fi*. www.head-fi.org/threads/headphone-measurements-the-new-standard-part-1.937301/—post-15735920 ⏎

Matteo, T. and Truchet, L. (2023). “Playing with Others Using Headphones: Musicians Prefer Binaural Audio with Head Tracking over Stereo”, *IEEE*, vol. 54, no. 3, pp. 501–511. ⏎

McAnally, K. I. and Martin, R. I. (2002). “Variability in the Headphone-to-Ear-Canal Transfer Function”, *J. Audio Eng. Soc*., vol. 50, no. 4, pp. 263–266. ⏎

Miller, J. (2024). “Sphere of Influence: How Harman and Dr. Sean Olive Are Reshaping Headphone Sound”. www.crutchfield.com/S-y8UEL5H9txe/learn/crutchfield-visits-harman.html ⏎

Miller, T. and Downey, C. (2023). “Listener Preferences for High-Frequency Response of Insert Headphones”, *J. Audio Eng. Soc*., vol. 71, no. 10, pp. 707–718. https://doi.org/10.17743/jaes.2022.0094. ⏎

Møller, H., Hammershøi, D., Jensen, C. B., Sørensen, M. F. (1995a). “Transfer Characteristics of Headphones Measured on Human Ears”, *J.*

Audio Eng., vol. 43, no. 4, pp. 203–217. ⏎

Møller, H., Jensen, C. B., Hammershøi, D. and Sørensen, M. F. (1995b). "Design Criteria for Headphones", *J. Audio Eng. Soc.*, vol. 43, pp. 218–232. ⏎

Møller, H., Sørensen, M. F., Jensen, C. B. and Hammershøi, D. (1996). "Binaural Technique: Do We Need Individual Recordings?", *J. Audio Eng. Soc*., vol. 44, no. 6, pp. 451–469. ⏎

Muchnik, C. Amir, N., Shabtai, E., Kaplan-Neeman, R. (2012). "Preferred Listening Levels of Personal Listening Devices in Young Teenagers: Self-reports and Physical Measurements", *J. Audiol*., vol. 51, no. 4, pp. 287–293. https://pubmed.ncbi.nlm.nih.gov/22122401/ ⏎

Olive, S. E. (2004). "A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 2—Development of the Model", *117th Convention Audio Eng. Soc*., Paper 6190. ⏎

Olive, S. E. (2009). "A New Reference Listening Room for Consumer, Professional and Automotive Audio Research", *126th Convention Audio Eng. Soc*., Paper 7677. ⏎

Olive, S. E. (2024). "A Comparison of in-ear Headphone Target Curves for the Brüel & Kjær Head & Torso Simulator Type 5128", *157th Convention Audio Eng. Soc*., Paper 10193. ⏎

Olive, S. E. and Welti, T. (2012). "The Relationship between Perception and Measurement of Headphone Sound Quality", *133rd Convention Audio Eng*., Paper 8744. ⏎

Olive, S. E. and Welti, T. (2015). "Factors That Influence Listeners' Preferred Bass and Treble Balance in Headphones", *139th Convention Audio Eng. Soc*., Paper 9382. www.aes.org/e-lib/online/browse.cfm?elib=17940. ⏎

Olive, S. E., Welti, T. and Khonsaripour, O. (2016a). "The Preferred Low Frequency Response of In-ear Headphones", *Int. Audio Eng. Soc. Headphone Conference*, Alborg, Denmark. ⏎

Olive, S. E., Welti, T. and Khonsaripour, O. (2016b). "Validation of a Virtual In-Ear Headphone Listening Test Method", *141st Convention*

Audio Eng. Soc., Paper 9658l. https://aes2.org/publications/elibrary-page/?id=18462 ⏎

Olive, S. E., Welti, T. and Khonsaripour, O. (2017a). "A Statistical Model That Predicts Listeners' Preference Ratings of In-Ear Headphones: Part 1 —Listening Test Results and Acoustic Measurements", *143rd Audio Eng. Soc. Convention*, Paper 9840. ⏎

Olive, S. E., Welti, T. and Khonsaripour, O. (2017b). "A Statistical Model That Predicts Listeners' Preference Ratings of in-ear Headphones: Part 2 —Development and Validation of the Model", *143rd Audio Eng. Soc. Convention*, Paper 9878. https://aes2.org/publications/elibrary-page/?id=19275 ⏎

Olive, S. E., Welti, T. and Khonsaripour, O. (2018a). "A Statistical Model That Predicts Listeners' Preference Ratings of Around-ear and On-ear Headphones", *144th Audio Eng. Soc. Convention*. www.aes.org/e-lib/browse.cfm?elib=19436. ⏎

Olive, S. E., Welti, T. and Khonsaripour, O. (2018b). "A Survey and Analysis of Consumer and Professional Headphones Based on Their Objective and Subjective Performances", *145th Convention Audio Eng. Soc*., Paper 10048. www.aes.org/e-lib/browse.cfm?elib=19774. ⏎

Olive, S. E., Welti, T. and Khonsaripour, O. (2019). "Segmentation of Listeners Based on Their Preferred Headphone Sound Quality Profiles", *146th Audio Eng. Soc. Convention*, Paper 1 ⏎

Olive, S. E., Welti, T. and McMullin, E. (2013a). "Listener Preference for Different Headphone Target Response Curves", *134th Audio Eng. Soc. Convention*, Paper 8867. www.aes.org/e-lib/browse.cfm?elib=16768 ⏎

Olive, S. E., Welti, T. and McMullin, E. (2013b). "Listener Preferences for In-Room Loudspeaker and Headphone Target Responses", *135th Audio Eng. Soc. Convention*, Paper 8994. ⏎

Olive, S. E., Welti, T. and McMullin, E. (2013c). "A Virtual Headphone Listening Test Methodology", *Audio Eng. Soc. 51st Int. Conference: Loudspeakers and Headphones*. www.aes.org/e-lib/browse.cfm?elib=16874 ⏎

Olive, S. E., Welti, T. and McMullin, E. (2014). “The Influence of Listeners’ Experience, Age, and Culture on Headphone Sound Quality Preferences”, *137th Convention Audio Eng. Soc.*, Paper 9177. www.aes.org/e-lib/online/browse.cfm?elib=17500. ⏎

Pralong, D. and Carlile, S. (1996). “The Role of Individualized Headphone Calibration for the Generation of High Fidelity Virtual Auditory Space”, *J. Acoust. Soc. Am.*, vol. 100, no. 6, pp. 3785–3793. https://doi.org/10.1121/1.417337. ⏎

Puro Sound Labs (2024). “The Puro Story”. https://purosound.com/pages/about-puro-sound-labs ⏎

Ravizza, G., Villegas, J., Volk, C. and Stegenborg-Andersen, T. (2023). ” Over-Ear Headphone Target Curve for Brüel & Kjær Head and Torso Simulator Type 5128 Measurements”, *155th Audio Eng. Soc. Convention*. ⏎

Rtings (2024). “The 6 Best Noise Cancelling Headphones—Fall 2024”. www.rtings.com/headphones/tests/isolation/noise-isolation-cancellation-passive-active ⏎

Rudmose, W. (1982). “The Case of the Missing 6 dB”, *J. Acoust. Soc. Am*. 71, pp. 650–659. ⏎

Sank, J. R. (1980). “Improved Real-ear Tests for Stereophones”, *J. Audio Eng. Soc.*, vol. 28, pp. 206–218. www.aes.org/e-lib/browse.cfm?elib=3994. ⏎

Scarrott, B. (2023). “Bluetooth Auracast Explained: 3 Big Reasons to Upgrade Your Wireless Headphones”. www.techradar.com/features/bluetooth-auracast-explained ⏎

Schäfer, M., Kopper, B., Reimes, J. and Gierlich, H. (2019). “Joint Reproduction of Background Noise and Reverberation for Development and Testing of Binaural Devices”, *Audio Eng. Soc. International Conference on Headphone Technology*. ⏎

Schärer, Z. and Lindau, A. (2009). “Evaluation of Equalization Methods for Binaural Signals”, *126th Convention Audio Eng. Soc.*, Paper 7721. ⏎

SenseLab (2024). “SenseLab Quality Rating Report: Premium True Wireless EarBuds 2.0”.

https://forcetechnology.com/en/services/acoustics-noise-sound-quality/senselab-quality-rating-reports/senselab-quality-rating-reports-premium-earbuds/senselab-quality-rating-report-on-premium-earbuds-2 ⏎

Silver, G. (2024a). "Soundstage Is (Much) More Complicated Than You Think". https://headphones.com/blogs/features/soundstage-is-much-more-complicated-than-you-think ⏎

Silver, G. (2024b). "Diffuse Field: Calculate, Characterize, Calibrate". https://headphones.com/blogs/features/diffuse-field ⏎

Simon, G., Olive, S. and Welti, T. (2009). "The Effect of Whole-Body Vibration on Preferred Bass Equalization in Automotive Audio Systems", *127th Convention Audio Eng. Soc*., Paper 7956. ⏎

Spatial Audio for Domestic Interactive Entertainment (SADIE) (2024). www.york.ac.uk/sadie-project/

Statistica (2023). "Digital Audio Usage by Device in the U.S. as of December 2023". www.statista.com/forecasts/997213/digital-audio-usage-by-device-in-the-us ⏎

Temme, S., Olive, S. E., Tatarunis, S., Welti, T. and McMullin, E. (2014). "The Correlation between Distortion Audibility and Listener Preference in Headphones", *137th Audio Eng. Soc. Convention*. www.aes.org/e-lib/online/browse.cfm?elib=17441. ⏎

Theile, G. (1986). "On the Standardization of the Frequency Response of High-quality Studio Headphones", *J. Audio Eng. Soc*., vol. 34, pp. 956–969. www.aes.org/e-lib/browse.cfm?elib=5233. ⏎

Toole, F. E. (1984). "The Acoustics and Psychoacoustics of Headphones", *Audio Eng. Soc. 2nd International Conference*, Paper C1006. ⏎

Toole, F. E. and Olive, S. E. (1988). "The Modification of Timbre by Resonances: Perception and Measurement", *J, Audio Eng. Soc*., vol. 36, no. 3, pp. 122–142. ⏎

Vafaei, S. (2018). "Raw Frequency Response", *Rtings.com*. https://bit.ly/3oBUnKG. Accessed November 2, 2021. ⏎

Völk, F. (2014). "Inter- and Intra-Individual Variability in the Blocked Auditory Canal Transfer Functions of Three Circum-Aural Headphones",

J. Audio Eng. Soc., vol. 62, no. 5, pp. 315–323. ⏎

Völk, F. and Fastl, H. (2011). "Locating the Missing 6 dB by Loudness Calibration of Binaural Synthesis", *131st Audio Eng. Soc. Convention*, Paper 8488. ⏎

Welti, T. (2015). "Improved Measurement of Leakage Effects for Circum-Aural and Supra-Aural Headphones", *138thAudio Eng. Soc. Convention*, Paper 9275. ⏎

Wykes, A. J. (2024a). "Headphone Driver Types: Which Ones's Just Right?". www.soundguys.com/driver-types-19347/ ⏎

Wykes, A. J. (2024b). "What Is MDAQS? The Multi-Dimensional Audio Quality Score Explained". www.soundguys.com/what-is-mdaqs-104949/ ⏎

CHAPTE
R 16

Hearing Loss and Hearing Conservation

Floyd Toole

DOI: 10.4324/9781003477495-17

This entire book is irrelevant if one cannot hear. It is only somewhat relevant if one is hearing-impaired. Good hearing is important to our pleasure, but if one is professionally involved with the audio industry, it is essential to being able to do our jobs. Dr. Brian C.J. Moore is well known for his informative and insightful research and writings on the physiology and psychology of hearing. The abstract of Moore (2016) states:

> Exposure to high-level music produces several physiological changes in the auditory system that lead to a variety of perceptual effects. Damage to the outer hair cells within the cochlea leads to a loss of sensitivity to weak sounds, loudness recruitment (a more rapid than normal growth of loudness with increasing sound level), and reduced frequency selectivity. Damage to the inner hair cells and/or synapses leads to degeneration of neurons in the auditory nerve and to a reduced flow of information to the brain. This leads to poorer auditory discrimination and

> may contribute to reduced sensitivity to the temporal fine structure of sounds and to poor pitch perception.

To any person interested in music, sound recording or sound reproduction, this is an alarming summary.

For audio professionals hearing loss is an occupational hazard, especially for those who also play in bands or have noisy hobbies. Anyone at work or at play who is exposed to loud sounds for prolonged periods, or to a single *very* loud sound, experiences some amount of hearing loss.

It is well known that temporary hearing loss results from exposure to any very loud sound. A ringing in the ears is the warning. With adequate quiet rest periods, we appear to recover, but if the exposures are repeated, some of the temporary loss becomes permanent. Whether the hearing loss is temporary or permanent it affects something as basic as one's ability to detect equalization changes (Kruk and Kin, 2015).

Section 10.1.2 shows that hearing loss well within the range of "normal" by standard audiometric criteria is sufficient to degrade one's ability to form good and consistent opinions about sound quality. People would listen carefully to a recording through a loudspeaker, form an opinion, and write a detailed explanation about what they heard. Then a few minutes or hours later they would hear exactly the same sound, form a very different opinion, and write a different essay. These were professional recording engineers, whose opinions determine what recordings sound like.

> *Audiometry focuses on speech intelligibility, not on hearing small nuances in music and movies. So, what is "normal" to an audiologist can be significantly impaired to an audiophile or audio professional.*

Currently, the fact is that once the inner ears have been damaged, the damage is permanent. Some audio information fails to reach the brain, and some that is delivered is corrupted. It is more than just not hearing small sounds; some higher-level aspects of perception occurring within the brain

are modified, even when threshold shifts are not present. It is a truly insidious affliction.

Clearly, it is important to do everything possible to prevent hearing loss, which is difficult as it has many causes:

- A single exposure to an extremely loud sound, such as an explosion, firing guns and many military weapons without protection, or automobile airbag deployment.
- Multiple exposures to very loud sounds, especially if sustained or repeated over a period of time that does not offer adequate quiet periods for recovery. The OSHA, NIOSH and other occupational hearing conservation criteria were developed from hearing loss measurements on multiple generations of factory workers in textile mills where the noise was constant for many decades.
- A gradual accumulation of occasional exposures to sounds that would be considered "normal" in daily life, work and recreation. Figure 9.3 shows a few example sound levels that we experience in life, and many of them are above the 75 dB SPL level that is believed to be the region above which permanent damage begins to accumulate.
- Over a lifetime one may experience any or all of the above, and this accumulated hearing loss is called presbycusis—the statistical normal hearing loss as a function of age. When an audiologist tells you that your hearing is normal "for your age", you are basically being told that your hearing is not what it used to be.
- Ototoxic (hearing damaging) drugs, including both prescription and over-the-counter medications. An internet search will reveal many, including some that may be in your medicine chest.
- Diving, disease, head trauma, surgical procedures, etc. There are also unexplained sudden onset losses, unilateral and bilateral.

Some of these may be avoidable, but not all. Critical listening is for those with close to normal hearing—within 15 to 20 dB of audiometric zero is a widely used criterion for listeners in double-blind tests (Section 10.1.2).

Now I live with a background of tinnitus. In my case it is a high-pitched hissing sound, almost a whistle. It varies in level depending on factors I have not been able to identify, but it is always there. I also am more sensitive to loud sounds than when I was younger. It is called hyperacusis (www.hyperacusis.net). I know people for whom even reasonably loud sounds inflict pain. I still enjoy dynamic range in movies and music, but not with the same enthusiasm as I used to. I occasionally hear distortion that is not in the sound arriving at my ears.

All this happened in spite of the fact that I have been very careful about wearing ear protection, starting in the 1970s. The hard-shell ear defender with liquid-filled cushions had been invented by two of my NRCC colleagues. They provided persuasive arguments about hearing protection. I wore these devices faithfully while mowing grass, operating power tools in my wood shop, and the like. I wore earplugs when driving my noisy (but enormously entertaining) sports cars.

The appearance of musicians' earplugs, a development of Dr. Mead Killion at Etymotic Research, revolutionized my hearing protection program (www.etymotic.com). These are custom-fitted earmold devices that reduce the volume without destroying timbre. The 15 dB attenuating version is useful. I remain in touch with my surroundings, headphones work on top of them, and I can sleep comfortably. They fit into my shirt pocket. At the end of a long flight, when I take them out I realize just how much noise I had been avoiding. They are compulsory at loud concerts, where one notices that the lyrics are actually intelligible, but the bass still shakes the torso. Now, several companies make musicians' earplugs. All of them reduce sound levels, and they all change timbre, some more than others (Anderson, 2023).

Now, miniaturization has allowed insert headphones to perform several useful functions in addition to delivering music and movie soundtracks. Reducing background noise is a convenient one. Limiting the maximum

playback sound level is an essential feature, because many headphones can deliver maximum sound pressure levels in the vicinity of 110 to 120 dB SPL, and there is evidence of elevated hearing loss among young people as a result. Transparency mode, allowing one to hear what is going on outside, is another. But hearing what is happening outside, while limiting the maximum sound level that reaches the ears, keeping it at safe levels, is the feature of interest in this context. Even hearing threshold measurements are possible; accuracy to be confirmed. And, finally, such smart headphones can function as hearing aids for moderate hearing loss. It is a rapidly changing world, for the better in this respect.

16.1 OCCUPATIONAL NOISE EXPOSURE LIMITS

I will begin by getting something off my chest.

> *Occupational hearing conservation programs are almost totally irrelevant to audio professionals and serious audiophiles.*

It is a bold assertion, but it is supportable. Figure 9.3 contains the OSHA occupational noise exposure limits. They suggest that a sound level of 90 dBA is acceptable for an 8-hour work day—that is all day, 5 days a week, every working week. But we know that anything above about 75 dB can have progressively greater effects on hearing. What is going on here?

It is essential to know that in existing national and international standards the *only* criterion being considered is the preservation of the ability to understand speech. The OSHA, NIOSH and similar occupational noise exposure criteria were created for manufacturing and other industrial workers and the goal was not to *prevent* hearing loss, it was to *preserve* enough that at the end of a working life, conversational speech at 1 m distance was possible. Permanent damage of an important kind is inevitable. Hi-fi hearing, critical listening ability is not preserved.

Distilled to its essence, it is considered acceptable for hearing loss to accumulate up to 25 dB in both ears at the 1 kHz, 2 kHz and 3 kHz

audiometric frequencies. In practical terms this translates to a loss of about 10% understanding of entire sentences, about 50% misunderstanding of monosyllabic "PB" words (words that are ambiguous because of similar sounding consonants) during conversation at normal voice levels, in the quiet, with persons 1 m apart (Kryter, 1973). And this is considered to be an acceptable situation: "normal" hearing. Further losses from 25 to 40 dB are described as "slight". Really? For whom?

16.1.1 What Can Audio Professionals Do for Protection?

In most noisy occupations, hearing protectors can be worn without affecting how workers do their jobs. Recording engineers, front-of-house engineers, and those working with or around these people cannot avoid being exposed to loud sounds at times. It is part of doing their jobs. Professionals know that their hearing is at risk, and many do what they can to minimize extreme exposures. Some wear musicians' earplugs, supposedly providing attenuation with minimal change of timbre. Anderson (2023) describes measurements indicating that these devices are not acoustically neutral, so the timbre of what is heard is changed. This can be tolerated for certain tasks, but for others unimpeded ears are necessary, and this is where hearing is at risk.

Jamieson (2015) and Compton (2022) provide detailed accounts of the challenging situation audio professionals find themselves in. The problem is that damage can be accumulating in the inner ear hair cell structures long before measurable threshold elevation is apparent. Notions that experience can overcome hearing loss is true only to the extent that many tasks can become routine, and hearing only the basic elements can suffice some of the time. The problem is that all of the time not everything, good or bad, is heard.

For audio professionals it is an almost impossible balancing act but necessary to avoid the unpleasant option of hearing aids at some point in life. The ultimate goal is to satisfy the customer. Monitoring in more consumer-like circumstances at more consumer-like sound levels is an option. Home studios may be a good trend. It seems to me that if a

recording sounds good in a well-designed stereo listening or home-theater system, it has "translated" for what may be the most demanding audience. Good home theaters can deliver full audible bandwidth sound at cinema sound levels, but most people listen at much lower levels.

16.2 NON-OCCUPATIONAL NOISE EXPOSURE

Many people underestimate the high sound levels that accompany many leisure activities. Hearing damage from recreational noises is sometimes called "sociocusis". Figure 9.3 shows a few examples of dangerously high sound levels, but much more extensive data and analysis are in Clark (1991). There are many everyday sounds that reach dangerous levels, some around the home or home workshop, and numerous combustion-engine-powered machines. Of them all, though, Clark concludes that the most serious threat to hearing comes from recreational hunting or target shooting, where both ears experience loss, but the ear opposite to the one pressed against the gunstock (most often the left ear) can show thresholds that are higher by 15 to 30 dB.

Much more subtle is the effect of listening to headphones at excessively high sound levels. Section 15.3.3 addresses the topic, with evidence of premature hearing loss among young people.

16.3 HIDDEN HEARING LOSSES

We think of hearing loss as an inability to hear small sounds, especially at high frequencies, but this is just part of the story. The sounds we can hear are modified; our perceptions are changed. One aspect is that we lose frequency resolution in our perceptual spectral-analysis capabilities. Section 9.6.3 stated:

> With hearing loss, the critical bands get wider, and, as a result, even our perception of music is altered. It is not just a matter of hearing or not hearing; it affects how we perceive the sounds that we are able to hear.

Speech intelligibility also suffers (Leek and Molis, 2012).

Coupled with this is *hidden hearing loss*, which can exist with or without elevated audiometric thresholds. It involves reduction in binaural discrimination, which results in difficulty separating the components of an acoustically complex situation with multiple sound sources and/or reflections. Many of us recognize that carrying on conversations in restaurants gets more difficult with age and hearing loss. Movie mixers deal constantly with dialog competing with background music, sound effects and ambient sounds in the multichannel (i.e. multidirectional) sound field. Are the artistic decisions they make likely to be affected by this kind of hearing loss?

The disquieting aspect of this affliction is that it can exist with little or no measured elevation of the hearing threshold. Kujawa and Liberman (2009) concluded:

> It is sobering to consider that normal threshold sensitivity can mask ongoing and dramatic neural degeneration in noise-exposed ears, yet threshold sensitivity represents the gold standard for quantifying noise damage in humans. Federal exposure guidelines (OSHA, 1974; NIOSH, 1998) aim to protect against permanent threshold shifts, an approach that assumes that reversible threshold shifts are associated with benign levels of exposure. Moreover, lack of delayed threshold shifts after noise has been taken as evidence that delayed effects of noise do not occur. The present results contradict these fundamental assumptions by showing that reversibility of noise-induced threshold shifts masks progressive underlying neuropathology that likely has profound long-term consequences on auditory processing. The clear conclusion is that noise exposure is more dangerous than has been assumed.

Liberman et al. (2016) elaborates on the situation with evidence that college-age students can exhibit hidden hearing loss. Bernstein and Trahiotis (2016) conclude that "listeners whose high-frequency monaural hearing status would be classified audiometrically as being normal or 'slight loss' may exhibit substantial and perceptually meaningful losses of binaural processing". These are troubling findings.

Some additional recent results add to this picture. The tests involved evaluations of Spatial Release from Masking (SRM), which is a measure of the angular separation between a "target" voice and an interfering voice necessary for the listener to be able to spatially separate them. This relates to the ability to separate images in a soundstage, something of increased importance in immersive audio. Srinivasan et al. (2016) concluded that young-normal-hearing listeners could discriminate approximately 2°, older-normal-hearing listeners could discriminate approximately 6°, but older-hearing-impaired listeners found little advantage even at 30° separation. In the end, age is a factor, but hearing loss, regular and/or hidden, determines the ability to discriminate among multiple images in a panorama.

Audio professionals tend to seek simplicity in the acoustical sound fields in which they work, choosing reflection-reduced, or near-field, dominant direct sound circumstances. Is it because of deteriorated hearing? Possibly.

16.4 WHAT IS ENJOYMENT? IT MIGHT DEPEND ON SPEECH CLARITY

To achieve high scores in speech intelligibility, a signal-to-noise ratio of 5 dB is good, and 15–20 dB nearly perfect. Noise, in this context, is everything other than the speech we want to hear. When several people talk at the same time, the noise is speech itself, and the sound spectra are very similar, making discrimination more difficult. In movies, the biggest problems are in the multichannel programs. Sound mixers pay great attention to dialog, but they have advantages over the rest of us. They get to hear each section of a film many times as they develop the audio design. They may know the dialogue before they even hear it and understand it even if they don't pay attention.

The ability to rewind and play, and the options of having subtitles or closed captions exist for a reason. Within the same room and playback system, one can switch from highly intelligible "talking head" television programs to watch a movie, and dialogue is not always perfectly understood? Part of it is the dramatic effect of mumbles and whispers;

movies have a significant dynamic range, whereas TV news and documentaries are highly compressed—always loud. Part of it is the inability to pick up clues from the lips, facial expressions, etc. when the talker is not facing the camera. Part of it is the mixture of music and sound effects emerging from all of the loudspeakers in the room, creating atmosphere and supporting on-screen action. The result is the same; more "noise", less intelligible speech.

Shirley and Kendrick (2004) investigated the effects of differing amounts of "extra" sound on listener impressions of dialogue clarity, overall sound quality and enjoyment. Clarity is a quality that corresponds to intelligibility, although it is not a direct quantitative measure of that parameter. Some of their listeners had measurably normal hearing, while others had varying degrees of impairment.

The test conditions involved only the three front channels, L, C and R, replaying components of twenty 5.1-channel film clips having durations ranging from 1 to 1.5 minutes. The variable was the level of extraneous sounds delivered by the L and R loudspeakers, while the center loudspeaker delivered a constant level of dialogue. The first condition ran all three channels at reference level. The second condition attenuated the L and R channels by 3 dB, the third condition attenuated them by 6 dB, and the last condition ran the center channel alone.

The results shown in Figure 12.8 are interesting from several points of view; moving from left to right in the sequence of presentation styles results in a progressive increase in the dominance of the center channel. In terms of dialogue clarity, obviously *all* listeners thought this was a good idea, those with normal hearing as well as those with impaired hearing. This confirms that the structure of the soundtrack is a major factor in dialogue clarity and intelligibility (and remember, surround channels are not involved in this test). The hearing-impaired listeners never got to the high levels of "clarity" reported by the normal-hearing group, but they appreciated the improvement achieved by attenuating the L and R channels.

In terms of "overall sound quality", there is fundamental disagreement between the groups. Those with normal hearing preferred all of the

channels running at or close to reference levels; anything else is a degradation. In total contrast, listeners with hearing disabilities, having difficulty understanding speech, clearly voted for progressively more monophonic sound. They seemed to equate speech clarity with sound quality, which is understandable.

When we come to ratings of "enjoyment" the hearing-impaired again told us that if the dialogue is not clear, then the movie is not enjoyable. Those listeners with normal hearing had trouble with this category, because variations in the ratings were high, and differences between the averaged ratings shown here are not statistically reliable. To the extent that they have meaning, it is interesting that these listeners entered *any* votes for turning the L and R channels down. But, they did.

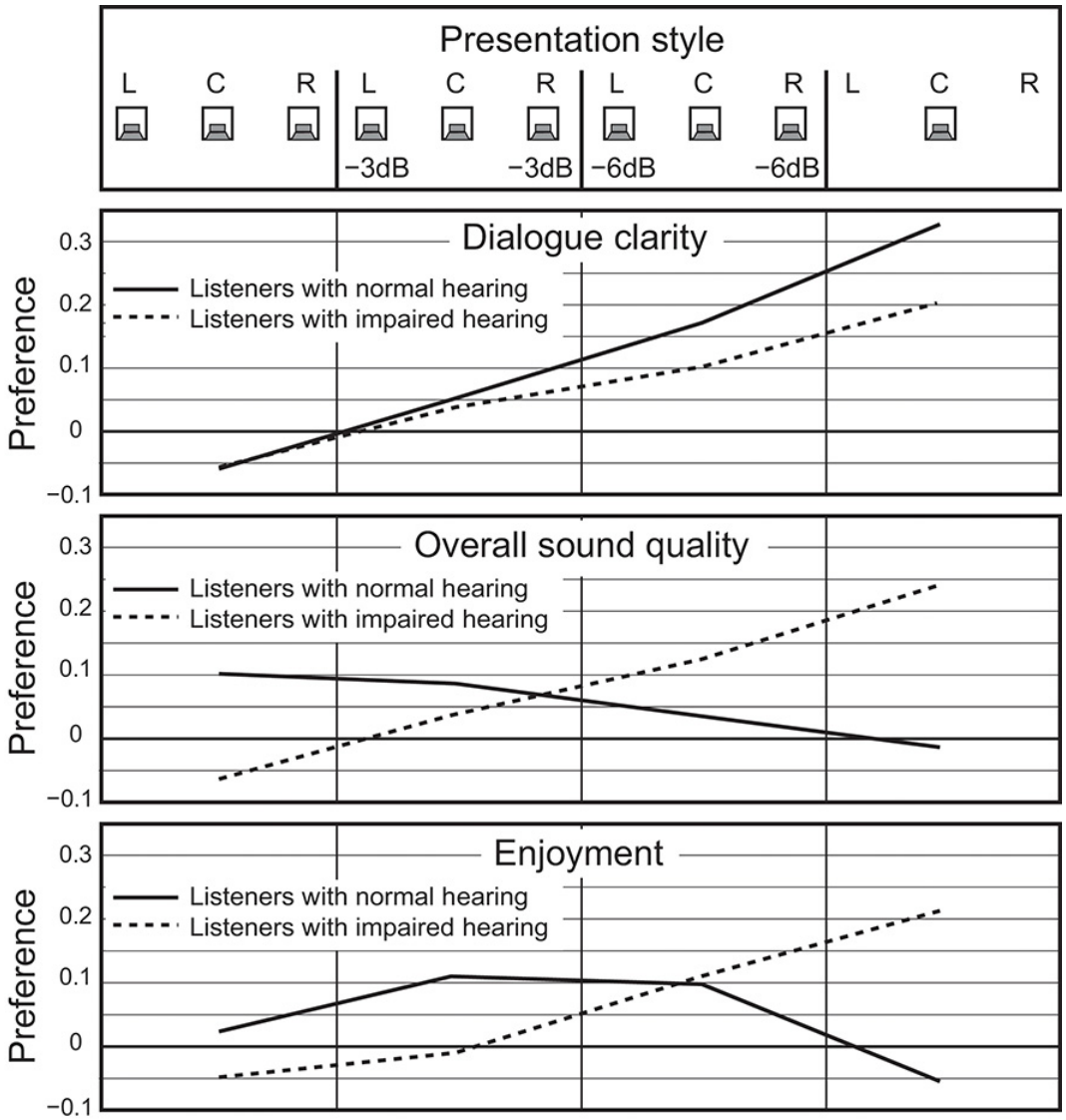


FIGURE 16.1 *The subjective preferences for various LCR front-soundstage presentations judged in the categories of "dialogue clarity", "overall sound quality" and "enjoyment". From left to right, the presentation styles progressively emphasize the center channel, ending with it in isolation. Listener responses were divided into two groups, listeners having normal hearing performance, and those having hearing impairments. Compiled from data in Shirley and Kendrick (2004).*

Listeners with impaired hearing were utterly predictable. Anything less than mono was degradation. This does not mean that they dislike the

experience of a multichannel presentation; it means that they place a higher priority on the clarity of dialogue.

These data do not paint an attractive picture for multichannel audio for those among us with deteriorated hearing. The good news is that other tests done by these authors indicate an improvement in speech clarity when the person speaking is facing the camera. Consciously or subconsciously, we all read lips. It is interesting to note that in their population of 41 subjects, with ages ranging beyond 75, those exhibiting hearing impairment began showing up in the 30–44 age group, with progressive deterioration at more advanced ages. The message here is that this is a factor to be considered in the creation of programs. The challenge is to decide what to do about it.

A common problem in the industry is how to deal with the perceptual difference between a real and a phantom center image. Three years after this work was done, Shirley et al. (2007) found that the interaural crosstalk in a double-mono phantom center image reduced speech intelligibility (see Section 4.3.2). Many multichannel programs avoid using the center channel, which does not help. I personally notice the difference when television channels deliver talking-head audio through a stereo pair, or a center channel. Ironically, some that avoid the center channel for their news programs use it and surround channels for commercials. Why? The improved clarity of a center channel voice delivery is obvious.

This is not news to the motion picture industry. Allen (2006) mentions examples of dialog intelligibility problems traceable to both the director of the film and the operators of the theaters. If film soundtracks are too loud, they get turned down, and dialogue levels drop along with those of the special effects.

Adjusting the center-channel level balance is a useful feature, and some systems allow for it. In the meantime, there are subtitles, the reason many prefer watching at home.

16.5 OBSESSION CAN BE A GOOD THING

Most of us know people who rely on hearing aids. I know several, and not one of them is pleased with the result. Hearing is not restored. It is more functional than without the devices, but the true essences of the audible world are gone forever. I have some hearing loss and tinnitus, and I am obsessive about preserving what I have left. I have made it to 87 without hearing aids, but the future is unknown. My opinions about sound quality are my own.

Each of us is responsible for our own hearing, but in life unexpected things happen. Many of us have lived portions of our lives without being conscious of the real risks to hearing and the consequences of hearing loss. Occupational noise exposure limits are not sufficient for anyone who wants to enjoy the full panorama of music and movie sound, and to be able to carry on a conversation around a table in a crowded restaurant or at a party. We all need to be somewhat obsessive about avoiding pointless loud sounds so we can enjoy the rewarding ones.

REFERENCES

Allen, I. (2006). "The X-Curve: Its Origins and History", *SMPTE Mot. Imag. J.*, vol. 115, July/August, pp. 264–273. ⏎

Anderson, D. (2023). "Analysis of Musical Spectral Distortions as a Rating Mechanism for High-Fidelity Earplugs", *155th Convention Audio Eng. Soc*., Preprint 10664. ⏎

Bernstein, L. and Trahiotis, C. (2016). "Behavioral Manifestations of Audiometrically-defined 'Slight' or 'Hidden' Hearing Loss Revealed by Measure of Binaural Detection", *J. Acoust. Soc. Am*., vol. 140, pp. 3540–3548. ⏎

Clark, W. (1991). "Noise Exposure from Leisure Activities: A Review", *J. Acoust. Soc. Am*., vol. 90, pp. 175–181. ⏎

Compton, S. (2022). "Managing the Live-Sound Audio Engineer's Most Essential Critical Listening Tool", *J. Audio Eng. Soc*., vol. 70, no. 4, pp. 305–318. https://doi.org/10.17743/jaes.2021.0065 Open Access. ⏎

Jamieson, A. (2015). “Use of and Attitudes towards Hearing Protection in the Sound and Music Industries: Results of a Pilot Survey”, *Audio Eng. Soc., 58th International Conference*. ⏎

Kruk, B. and Kin, M. (2015). “Perception of Timbre Changes vs. Temporary Threshold Shift”, *138th Convention Audio Eng. Soc*., Preprint 9228. ⏎

Kryter, K. D. (1973). “Impairment to Hearing from Exposure to Noise”, *J. Acoust. Soc. Am*., vol. 53, p. 1211. ⏎

Kujawa, S. G. and Liberman, M. C. (2009). “Adding Insult to Injury: Cochlear Nerve Degeneration after ‘Temporary’ Noise-Induced Hearing Loss”, *J. Neurosci*., vol. 29, pp. 14077–14085. A summary can be found at: http://newswise.com/articles/noise-induced-hidden-hearing-loss-mechanism-discovered ⏎

Leek, M. and Molis, M. (2012). “Hearing Loss and Frequency Analysis of Complex Sounds”, *Acoust. Today*, vol. 8, no. 2, pp. 29–33. ⏎

Liberman, M., Epstein, M., Cleveland, S., Wang, H. and Maison, S. (2016). “Toward a Differential Diagnosis of Hidden Hearing Loss in Humans”, *PLoS ONE*, vol. 11, no. 9, p. e0162726. https://doi.org/10.1371/journal.pone.0162726 ⏎

Moore, B. C. J. (2016). “Effects of Sound-Induced Hearing Loss and Hearing Aids on the Perception of Music”, *J. Audio Eng. Soc*., vol. 64, pp. 112–123. ⏎

NIOSH Occupational Noise Exposure (1998). www.cdc.gov/niosh/docs/98-126/pdfs/98-126.pdf ⏎

OSHA Occupational Noise Exposure (1974). http://osha.gov/dts/osta/otm/new_noise/index.html ⏎

Shirley, B. G. and Kendrick, P. (2004). “ITC Clean Audio Project”, *116th Convention, Audio Eng. Soc*., Preprint 6027. ⏎

Shirley, B. G., Kendrick, P. and Churchill, C. (2007). “The Effect of Stereo Crosstalk on Intelligibility: Comparison of a Phantom Stereo Image and a Central Loudspeaker Source”, *J. Audio Eng. Soc*., vol. 55, pp. 852–863. ⏎

Srinivasan, N. K., Kasey, M. J. and Gallum, F. J. (2016). “Release from Masking for Small Spatial Separations: Effects of Age and Hearing Loss”, *J. Acoust. Soc. Am. Express Letters*. http://dx.doi.org/10.1121/1.4954386 ⏎

CHAPTE
R 17

Summary and Conclusions

Floyd Toole

DOI: 10.4324/9781003477495-18

When I took an interest in audio as a teenager, in the 1950s selecting a radio or loudspeaker relied on finding one that matched one's personal preference. "Good tone" was the goal. Unconfirmed opinions, sprinklings of facts, and exaggerated marketing stories defined the industry. A similar situation prevailed after graduation in 1966 when I planned to continue my sound localization research at the National Research Council of Canada. I needed technically accurate loudspeakers to deliver specific sounds to listeners in an anechoic chamber. When I measured a few highly regarded products I was astonished at how ragged the anechoic frequency response curves looked. I expected better. Could these loudspeakers sound as bad as they looked?

A few simple, blind, equal-loudness, multiple-loudspeaker comparison listening tests conducted on a Friday afternoon revealed that indeed some of them did sound as bad as the measurements looked. The sound qualities

varied enormously. But the subjective “fidelity” ratings yielded some surprises. First, most of the listeners agreed on which loudspeakers were preferred. Where were the personal preferences? Second, the loudspeakers that were preferred exhibited the smoothest anechoic frequency responses. Why did the bad ones exist?

It was these observations that initiated the research reported in this book. The explanations eventually became clear. Blind and double-blind subjective evaluations that were free from bias were virtually non-existent at the time. So, opinions reigned supreme. We realized that humans are very adaptable creatures and can learn to live with very flawed loudspeakers. However, when exposed to something better, they recognize the improvement.

On the measurement side, there was no agreement on what to measure. The methods in use covered all possibilities, from anechoic on axis, to sound power, to steady-state room curves. They cannot all be right. 1/3-octave resolution measurements were common, so much that we now know to be audible could not be seen. The belief that we cannot measure what we can hear was understandable at that time. Progress accelerated in the early 1980s, when digital technology provided the means to measure, store and post-process data into forms that visually communicated more information.

Applying scientific methods to improve the creation and delivery of the art in movies and music has been the goal. It is a challenging process, involving human listeners with their numerous capabilities and frailties, but in the end, listening is the final judge of sound quality. Fortunately, in the right circumstances, listeners can be remarkably reliable “measuring instruments”, generating repeatable opinions that can be treated as “facts”. Through these subjective insights researchers learned what measurements mean, which ones are most useful, and ultimately how to design sound reproducing devices, loudspeakers and headphones, that will appeal to listeners.

At the present time, an understanding of spinorama or comparable anechoic measurement data alone allows users to select loudspeakers that are essentially neutral reproducers.

17.1 INTERPRETING ANECHOIC DATA

Ever since acoustical measurements were possible, a flat on-axis frequency response has been the performance target for most loudspeaker designers. The logic was that the spectrum of the original captured sound should be faithfully reproduced. Research confirmed the merit in this and added smoothness and an absence of resonances to the requirements. Still more inspection revealed that it was important to maintain those attributes in the far off-axis sounds that generate strong early reflections in rooms, the second loudest sounds we hear. Figure 17.1 provides an interesting comparison of data from my research, separated by 30 years.

The reduced variations in Figure 17.1 indicate that loudspeakers have clearly improved. The performance objectives of their designers did not change but measurements, engineering methods and transducers improved. Listeners 30 years and 2800 miles (4500 km) apart in different countries all preferred spectrally balanced, timbrally neutral loudspeakers.

In Figure 17.1(b) the predicted room curves tilt downward because the frequency-dependent directivity of forward-firing loudspeakers results in progressively more sound energy being delivered to the room at lower frequencies. Adding to this trend, reverberation time normally increases at lower frequencies and very high frequencies are reduced by air attenuation. With forward-firing loudspeakers, it is not possible for both the direct sound and the steady-state curves to be flat except in an anechoic space. The physical reality of bass rising in the reflected sounds following the direct sound, exists in most live listening experiences as well as in reproduced sound (see Figure 12.9).

It is easy to believe that steady-state room curves are indicators of sound quality because all the sound we hear is in them. What is missing is the recognition that two ears and a brain extract complex timbral, directional and spatial information from that sound. Special importance is given to the first sound arrival for localization and sound quality, and reflections are used to inform us about the listening environment. A microphone simply adds direct and all reflected sounds together, making it mono. Reflections

that are perceived as natural room sounds, spaciousness, are measured as comb filtering, something to be “fixed”. The problem is that by equalizing these audibly unimportant irregularities in room curves one is changing the very important direct sound. If reflections are a problem, acoustical treatment is the solution.

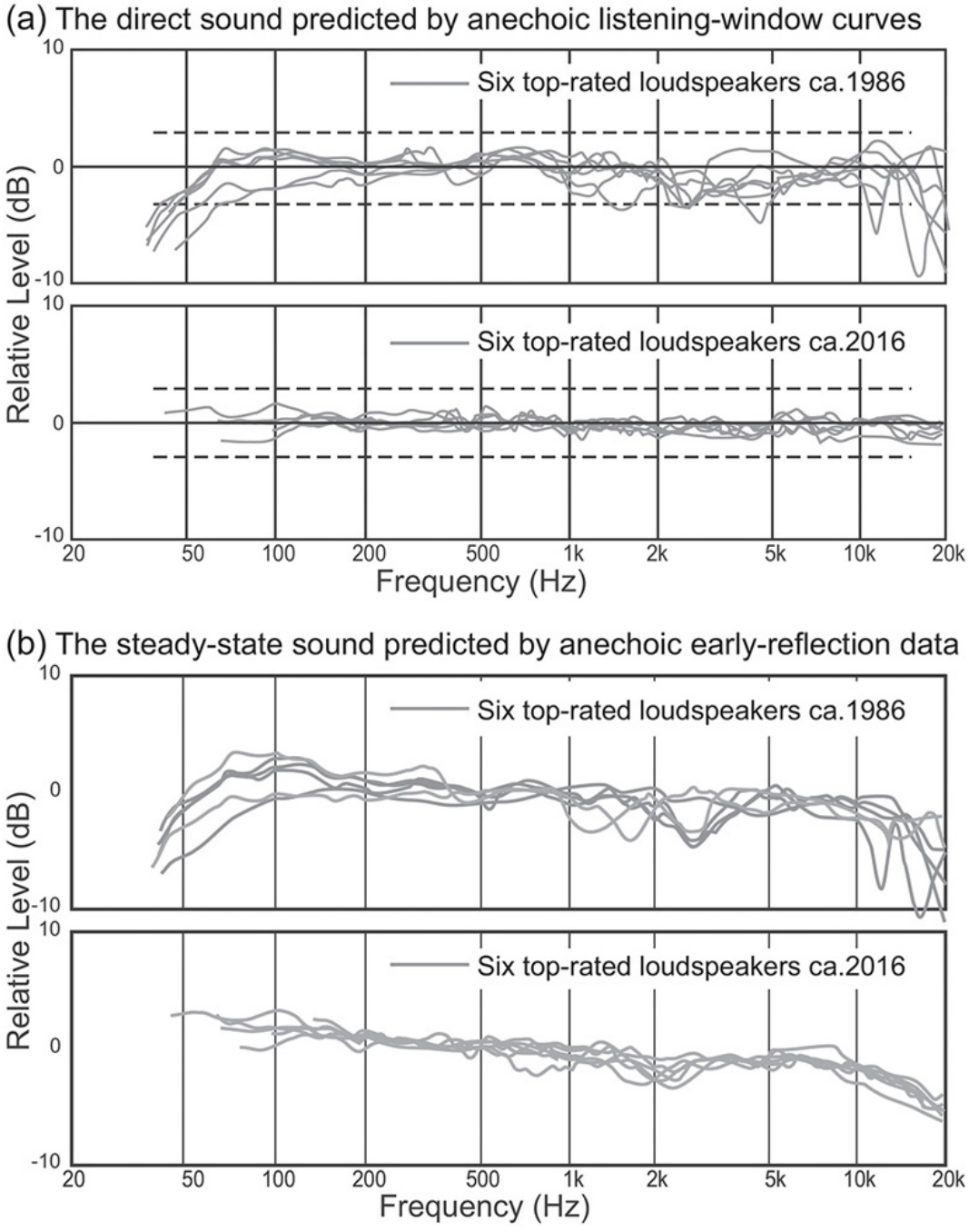


FIGURE 17.1 *Thirty years of improvement. A comparison of measured performances of six passive, unequalized loudspeakers that had the highest sound quality ratings in double-blind listening tests done in 1986 and 2016. (a) shows 1/20-octave listening window curves, with the traditional ± 3 dB tolerance window indicated by dashed lines. (b) shows the early-reflection curves, which are good predictions of steady-state room curves above the transition/Schroeder frequency.* ⏎

Now "room equalization/calibration" processes are businesses, claiming to adjust unknown loudspeakers, in unknown rooms to sound like "good" neutral loudspeakers. At low frequencies there may be truth in the claim, but above the transition/Schroeder frequency the situation changes. Without knowing the comprehensive anechoic performance of the loudspeaker, it is not possible to accurately interpret the meaning of a steady-state room curve to decide what can be corrected, how it should be corrected, and what should be ignored.

From double-blind tests reported in 1986, listeners have indicated that the direct sound, the on-axis frequency response, is the dominant factor in deciding perceived sound quality in rooms. It has been verified in different ways since, as described in this book. The room curves are *results*, not *targets* because they are dominated by off-axis sound radiation, and loudspeakers differ in their frequency-dependent directivities. Only at short listening distances or in very dead rooms would the room curve be similar to the on-axis or listening window curve.

In normal listening situations, equalizing a flawed loudspeaker to have the same room curve as a well-designed one does not ensure equivalent sound quality *except* at low frequencies, where all loudspeakers are essentially omnidirectional.

Figure 17.2 shows examples of four loudspeakers that received high sound quality ratings in double-blind evaluations. They all have flat and smooth on-axis and listening window, i.e. direct sound frequency responses. The anticipated steady-state room curves that result are well predicted by

the early-reflections curves at normal listening distances in typical listening rooms. These all tilt downward, but they are not targets, they are results for these specific loudspeakers. What we hear below the transition/Schroeder frequency is best represented by the steady-state room curve that is measured at the prime listening location in the listening room after installation. The loudspeaker provides the low-frequency energy, but the room resonances and adjacent boundary effects determine the quality of the bass and which listening locations will sound best.

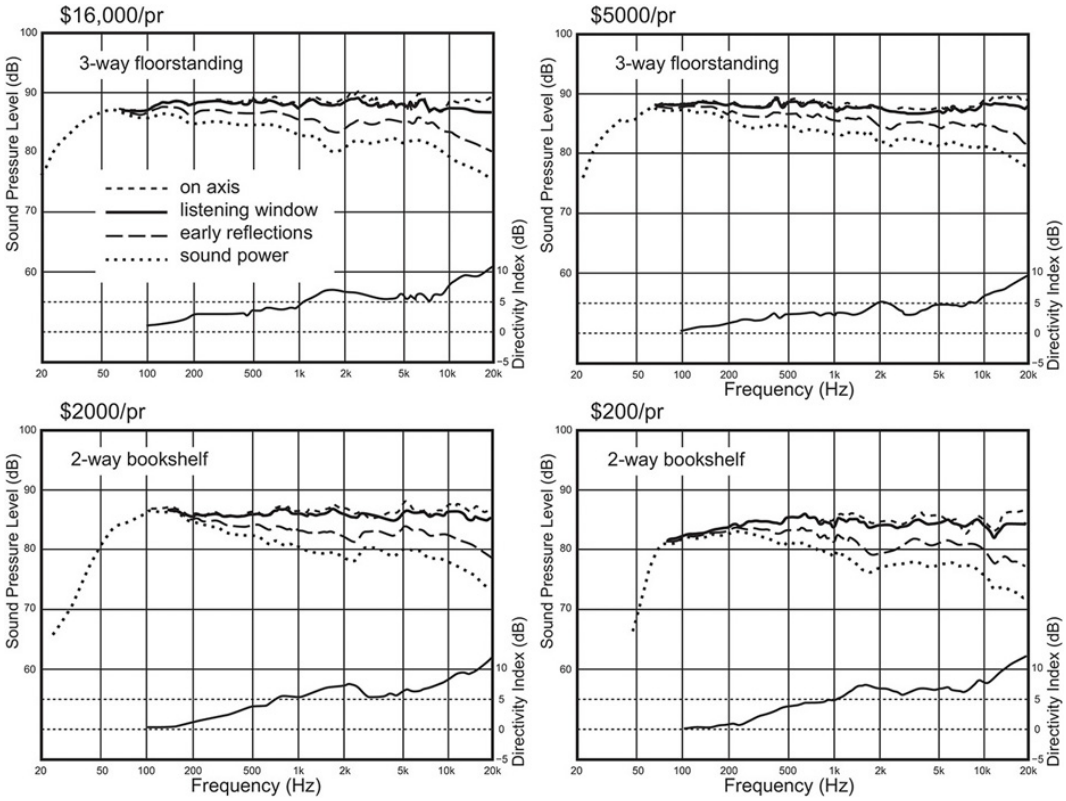


FIGURE 17.2 *Four passive forward-firing loudspeakers that received high subjective sound quality ratings. The prices are $200/pair and $2000/pair for bookshelf designs, and $5000/pair and $16,000/pair for floor-standing designs.*

The most important feature of these measurements is that there is no evidence of strong resonances. These would show up as repeated bumps in the family of curves because sound from resonant activity tends to radiate

in all directions, whereas acoustical interference bumps and dips change with angle and are averaged out in the spatial averaged curves below the on-axis curve. This characteristic combined with a flat direct sound is the prime indicator of a timbrally neutral loudspeaker. The shape or tilt of the off-axis curves is less important than their smoothness in this respect.

The bookshelf loudspeakers obviously need subwoofer assistance at low frequencies, but all the loudspeakers show commendable performance above about 80 Hz. The most obvious differences are in visual aesthetics, sound output capability and distortion. Price correlates with these factors, but not with basic timbral accuracy at moderate listening levels. There are also subtle differences in directivity, which will affect the interaction with adjacent boundaries in listening rooms, and the resulting spatial perceptions. However, in terms of basic sound quality these are all good products, and the $200/pair bookshelf unit is an excellent entry point for an aspiring audiophile.

Readers willing to explore the sources of measurements mentioned in Section 2.9 will find more products as good or better than these. There is no doubt that loudspeakers have improved in recent years. Increasingly, manufacturers are aiming for neutral timbre, but not all have equal skill in achieving it.

A non-performance factor adding to the cost of *some* higher priced loudspeakers is quality control. It is one thing to see an impressive set of measurements, but will they be replicated in the loudspeakers you buy? It is expensive to repair or scrap loudspeakers that don't pass a test at the end of the production line.

17.2 LOOKING BACK 60 YEARS

Figure 17.3 shows products from the 1960s. Some of those loudspeakers were aggressively colored, but their owners had enjoyed them until they heard something better in the blind listening tests. The loudspeakers that were evaluated were brought in by local audio stores or audiophiles, submitted for testing as a basis for reviews in Canadian audio magazines or

submitted by manufacturers. Many interesting products were missed. There was no budget to purchase loudspeakers in a government research laboratory.

Figures 17.3 and 17.4 show samples from products measured over the years. Several of these products performed well in double-blind listening tests at the time, but some had problems. A few showed evidence of skillful engineering. They came from the United States, United Kingdom, Canada and Japan, indicating that all regions made some decent sounding loudspeakers. I have not used valuable space to show poorly designed loudspeakers. They incorporated almost every conceivable flaw and several displayed respected brand names. Price was not a reliable indicator of sound quality.

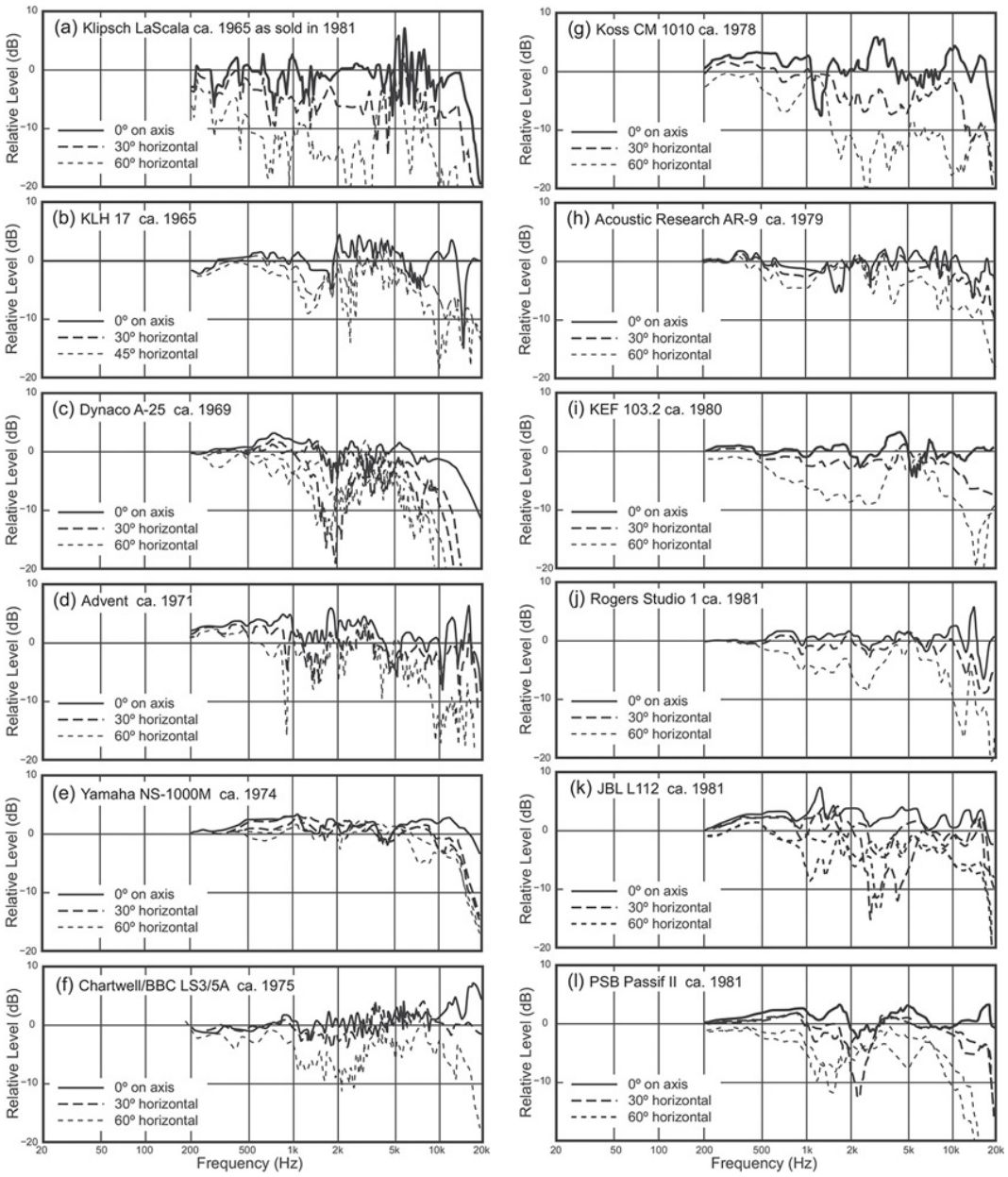


FIGURE 17.3 *Twelve consumer loudspeakers from the period 1965 to 1981. Those with asymmetrical driver arrangements were measured on both left and right sides. Curves truncated at 200 Hz were done using the first-generation analog measurement system that had errors at low frequencies. Sound levels have been normalized; relative sensitivity is not revealed.* ⏎

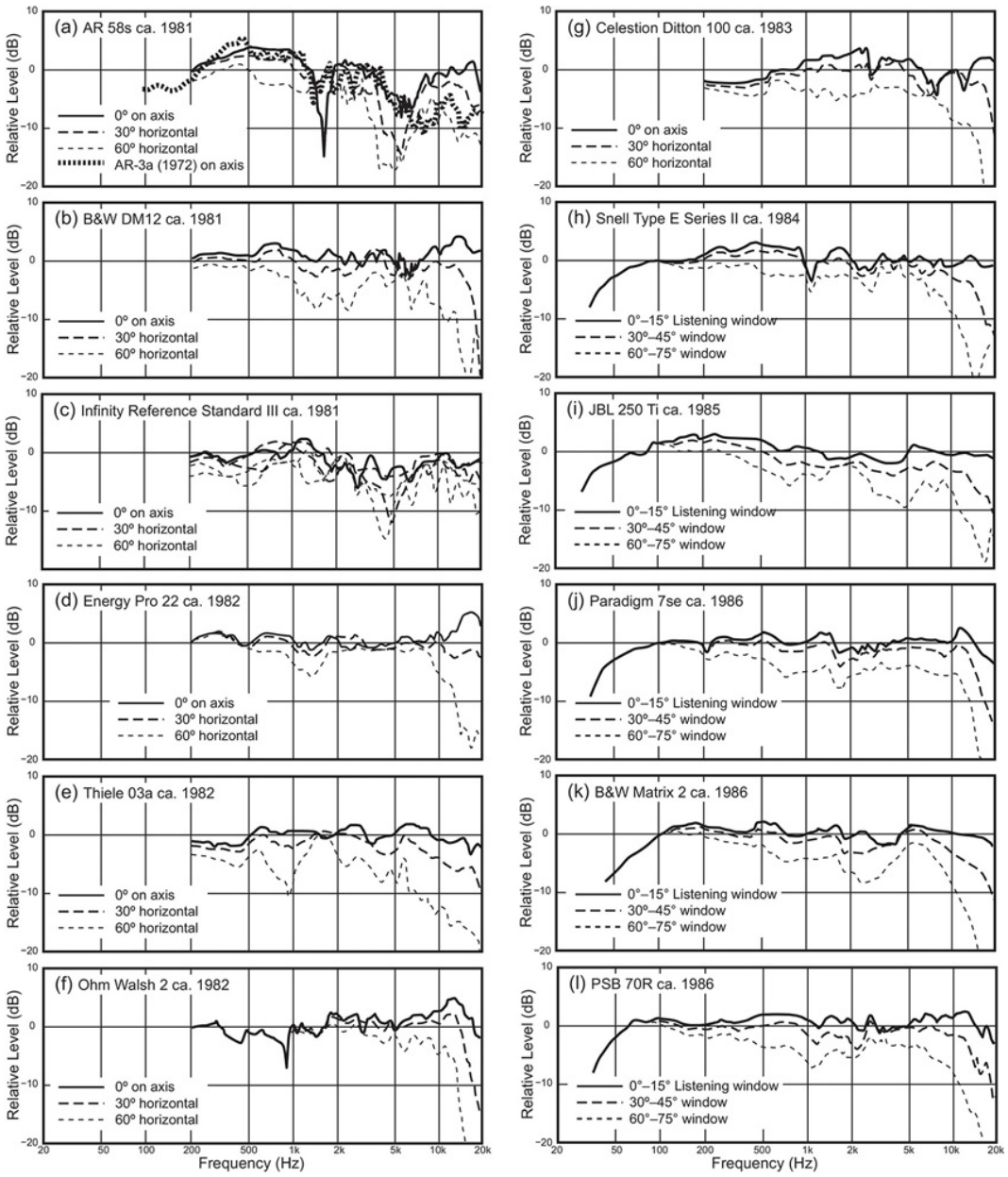


FIGURE 17.4 *Twelve loudspeakers that follow those in Figure 17.3, documenting progress in consumer loudspeaker design in the years 1981 to 1986.* ⏎

In Figure 17.3(a) a large horn loudspeaker exhibits evidence of acoustical interference and high directivity, but the underlying spectral balance was well maintained on and off axis. Very high sensitivity was a benefit at a time when amplifier power was limited. Loudspeaker (b) the Henry Kloss–designed KLH 17 and (c) the Danish Dynaco A-25 were popular and

affordable small loudspeakers of the period. Kloss also contributed the respected Advent (d) that, in a larger version, was often used in pairs—the "stacked Advents".

The Yamaha NS-1000M (e) was a serious effort, with beryllium dome midrange and tweeter units, mounted to minimize diffraction, to maximize and maintain constant directivity, and with obvious care in crossover design. They exhibited the lowest non-linear distortion of any consumer loudspeaker I had tested up to that point. The designer (also of the inferior NS-10M) visited me at the NRCC, modestly explaining that it was simply "good engineering". The curves show no evidence of resonances, which was the point of using beryllium. The design objective was a flat sound power curve for both the NS-1000M and the NS-10M, but that target was a better match for the almost constant directivity NS-1000M than it was for the conventional small two-way NS-10M, as illustrated in Figure 3.12. It was the wrong target, as the designer learned, but an example of excellent engineering.

The BBC contributed several loudspeaker designs over the years, and this Chartwell licensed version of the LS3/5A was the "outside broadcast monitor" intended for use in small dead spaces. It was a small two-way, using KEF drivers, much like the midrange/tweeter combination in floor-standing loudspeakers. It acquired an enthusiastic following. Samples from other licensed manufacturers have been different, some not very good—transducer variations or substitutions? The BBC connection was a persuasive selling feature.

Figure 17.3(h), the AR-9, was another serious effort, showing evidence of attention to both spectral balance and uniformity of directivity. It embodied boundary interaction control as advocated by Allison—see also Figure 13.11. It got high listener ratings.

The other loudspeakers in Figure 17.3—(g), (i), (j), (k) and (l)—all show that the designers aimed at flattish on-axis frequency responses. Some came closer than others, but all lost control in achieving uniformity in off-axis radiation. In normally reflective rooms, these spectrally unbalanced reflected sounds affect the experience.

Figure 17.4 continues the story, beginning with the AR 58s, which is dramatically different from the neutral spectral balance of the AR-9 in Figure 17.3(h). It was reminiscent of the AR-3 shown in Figure 1.3 and the very similar AR-3a, so the on-axis curve is superimposed for comparison. It looks like an attempt to revisit the sound of the old AR-3a. If so, it seems that the once desired AR-3a "improved" was achieved, because the tweeter now had more output, but the sound quality was definitely colored.

The DM12 had good on-axis response, but exhibited the classic 6-inch (150 mm) two-way off axis misbehavior. The Infinity (c) had good directional control but a mid-frequency hump determined the tonal balance of the product. The Energy (d) was uncommonly good in all respects: flat, smooth, almost constant directivity—it consistently scored well in listening tests. Others, as can easily be seen, were variations on a common theme. The Ohm Walsh 2 shown in (f) is an interesting design, being horizontally omnidirectional up to a crossover to a conventional tweeter. It was a simplification of the original single broadband driver invented by Lincoln Walsh.

Notably absent from this collection is the Quad ESL 63, but some data are shown in Figure 4.7. Also absent from the displayed products are those from Bose. To reveal useful performance data on their multidirectionally radiating loudspeakers requires the digital measurement system that can perform spatial averaging and a sound power calculation; it did not exist at the time I was doing those tests. Measurements on single axes show acoustical interference that is disturbing to the eye but not necessarily evident to listeners in normally reflective rooms. So, rather than show misleading data, I show none. The same applies to the image-stabilizing dbx Soundfield loudspeaker discussed in Section 4.3.3. These designs added interest to the panorama of products in the marketplace, appealing to some, not to others.

It is clear that as the years passed loudspeakers became increasingly similar. Products with very similar anechoic measurements are sometimes difficult to distinguish in double-blind tests. Often it comes down to a decision between sounds that are slightly different, but comparably good. In

those cases, program material become the deciding factor, favoring one or the other at different times, and the end result is a statistical tie. There is a point of diminishing returns. And, as pointed out, the listening room is a major factor at low frequencies.

This has been a perspective on consumer audio. Sections 3.5.1 and 3.5.2 provide comparable data on some professional monitor loudspeakers from the same period. Their problems affect the ability of recording engineers to hear accurately what is being recorded, and to estimate how well their product will translate to the outside world. It cannot have been easy. Figure 2.7 shows a compilation of double-blind *subjective* sound quality ratings on a mixture of consumer and professional loudspeakers indicating that consumers were doing very well indeed, even at moderate prices (and moderate sound levels).

17.3 WHERE ARE WE NOW?

There are two answers. One is based on hard data; see Figure 17.2 for a few examples. And then there is the answer based on opinions. The two can be remarkably, and depressingly, different. A persistent problem throughout the audio industry and consumer world is a lack of discipline in conducting subjective evaluations. Sighted evaluations are susceptible to bias, regardless of who the listener is, consumer or professional, inexperienced or experienced (see Section 10.1).

The “take it home and listen to it” process of subjective reviewing probably gratifies the reviewer, but it leads nowhere reliable. Reviews employing this method are as much a commentary on the recordings being auditioned as they are a critical analysis of loudspeaker performance—the variables cannot be reliably separated, and the room and its effect on bass is a significant uncontrolled variable unless the problems have been deliberately addressed. “Breaking in” is often mentioned and it is the listener that adapts, not the hardware that acquires musical abilities with time. It is significant that products never get worse. Without equal-loudness blind comparisons we get entertaining essays and opinions, and without

trustworthy measured data there is no way to understand or challenge the opinions. It is also unfortunate that reviewers usually do not reveal evidence of an important qualification for that job: normal hearing—ideally within 15–20 dB of audiometric zero. If not, their opinions are their own, not to be shared.

Loudspeaker reviewers who have access to trustworthy anechoic data have an enormous advantage. The headphone equivalent is measurements on an industry compliant test fixture. Their data are stable and meaningful references: facts. Double-blind comparative listening tests may not be possible, but by listening before measuring, and comparing what is heard to the measurements adds substance the opinions being offered. This is a welcome development for the industry, because well-designed products can be recognized and get credit. Readers also learn that measurements are important information.

Some loudspeakers are excellent out of the box, but there are many in the middle category: good in the basic essentials but needing a bit of equalization to really shine. Having anechoic data and a parametric equalizer allows the user to finish the job. This is what is done in active loudspeakers. Some reviewers helpfully provide suggestions. Equalization is almost always needed to address the low-frequency room resonance problems that we all share, so having equalization capability is a major benefit.

This situation is very different from that which existed only a few years ago. Meaningful measurements of loudspeakers and headphones have changed the industry. The evidence is in the products in the marketplace and in the way enthusiasts discuss them. Some audio forums frequently include measurements in the discussions. Mediocre sound and baseless marketing claims are still with us, but there is a clear trend towards timbrally neutral loudspeakers and headphones in both consumer and professional markets.

To gain better insight into what we base our opinions on, let us try to identify the key factors in what we hear.

- There is the *art*: the music, musical talent, and how they are blended and manipulated into a deliverable product by recording and mastering engineers.
- A large part of appreciating the art is *sound quality*: do the voices and instruments sound real, natural, if that was the intention? Abstract art can be colored, in which case customers should be able to hear what was intended.
- Another part of the art involves *direction and space*: the construction of images on a soundstage and the sense of spaciousness and envelopment that perceptually transport listeners to different venues.
- In a perfect world, all dimensions would "translate" to all listeners.

Loudspeakers and headphones are responsible for delivering sound quality. They should be neutral communicators—"honest messengers" of how voices and instruments sounded at the point of creation. In all listening modes, small-room resonances and adjacent-boundary effects are major influences on sound quality at frequencies below about 300–500 Hz. These must be addressed, as this portion of the spectrum accounts for approximately 30% of our overall impressions of sound quality. Most of the perceived directional and spatial dimensions are determined by the recording engineer, probably listening in a dominant direct sound field generated by conventional forward-firing loudspeakers.

These are the technical components behind the creation and delivery of the "art". Whether one likes the art itself is totally a personal, emotional thing. These are subjective/emotional perceptions, not measurable quantities. But there needs to be basic trust in the technical devices involved with the creation of the art and its delivery to consumers, and this requires a belief in measurements and of the science behind them. *Opinions totally dominate the artistic content, as they should, but if uncontrolled opinions are allowed to influence the technical portions of the operation, there cannot be a reliable basis for translation.*

The research discussed in this book attempts to bridge the cultural gap between subjective and objective, art and science, perspectives by providing evidence of reliable relationships between what can be measured and what is heard. What has been learned bears directly on the matter of translation.

Subjective evaluations done over the 60 years since this research was initiated have used commercial recordings as program material. Throughout, hundreds of listeners in hundreds of double-blind multiple-comparison tests have shown a remarkably reliable preference for the most timbrally neutral, resonance-free loudspeakers. The correlation between what is measured and what is heard is truly impressive (see Chapter 11).

It appears that some sound quality aspects of recordings "translated" very well indeed. In multiple-loudspeaker comparison evaluations listeners reliably separated the unwanted resonances in loudspeakers from the desirable ones in musical instruments and voices, all within the reflected sound fields of several different "typical" domestic listening rooms. Broadband bass and treble balance differences were present and recognizable in some of the recordings. These spectral balance variations are experienced in different programs, and some of them inevitably exist because of what was heard in recording and mastering rooms. They can detract from satisfaction when listening through neutral sound reproduction systems.

> *Being realistic, sound quality is the remaining and likely the dominant factor in deciding the successful communication, the translation, of the art. A persistent challenge is that bass is very important, and bass quality is dominated by the listening room and the setup within it.*

Loudspeakers and headphones are the prime determinants of perceived sound quality in reproduced sound. Their timbral neutrality is important because they represent—speak for—all musical instruments and vocalists, all pop and jazz groups, and symphony orchestras. Stereo and especially multichannel audio add important perceptions of envelopment, of being in large spaces, and immersion, of being surrounded by sounds from many directions. These physical, technical systems reproduce sound, and that

sound is art produced by musicians and crafted by recording engineers using these same systems. The art needs to translate from the creators to the customers. The circle of confusion should not exist. When good music recordings, movies and video concerts combine with good sound reproduction systems the experiences are truly remarkable entertainment.

Science in the service of Art

is our Business.

Let us get on with it.

Index

A vs. B paired comparison test method 16
absolute polarity 269, 272
ABX test method 17
acoustical absorbers: absorption in diffusers 380; angle-dependent absorption 375–377; bass traps/membrane absorbers 377–378; effect on audibility of reflections 185–186; porous/resistive absorbers 282–284; random incidence absorption coefficient 373–375
acoustical interference/comb filtering: audibility vs. measurement in rooms 183–187; definition 181–183
adaptation: in binaural hearing and HRTFs 74–77; related to "breaking in" 16; to rooms 132, 150–151
adjacent-boundary effects 400–402; boundary-friendly loudspeakers 408–410; in-wall and on-wall loudspeakers 231–234; listeners have boundary effects too 410–412; loudspeaker mounting options 403–408
air attenuation *see* propagation loss
Allan Devantier biography 55–56
Allison effect *see* adjacent-boundary effects
amplifier interactions with loudspeakers *see* damping factor
amplifier power requirements 248–250
ANSI/ASA S3.36 (2018) 480
ANSI/CTA-2034 loudspeaker measurement standard 48–49
apparent source width (ASW) *see* spatial effects
Apple Spatial Audio 472
Archie 488
around ear (AE) 476

array loudspeakers 234–237
audible effects summarized 131
audio industry described 67
auditory memory in listening tests 19
Auracast 474
Auro 3D *see* multichannel audio, immersive audio
automotive audio: advantages of multichannel 157; background noise 158, 260–261; typical frequency responses 157–158
A-weighting curve 320

bass in small rooms *see* room modes/resonances
bass management 217–219, 465–466
BBC monitors 47
Benade central paradox 25
binaural record/reproduction: through headphones 71–77; through loudspeakers 77–81; used in loudspeaker evaluations 142–144
bipole loudspeakers: floor standing (Mirage M1) 364; surround on-wall 231–232
blind and double-blind *see* listening tests
breaking in 16
B-weighting curve 320

car audio *see* automotive audio
CBC (Canadian Broadcasting Corporation) monitor tests 35, 47
CEDIA (Custom Electronics Design and Installation Association) 48
center channel loudspeakers 229–231
cinema sound: immersive systems 94; imperfect calibration standard – the X curve 90; improved calibration methods 91–93; incompatible with the rest of the audio industry 90, 93; inconsistent calibration results 91; introduction of surround sound 213; loudness concerns 239–240
circle of confusion in the recording industry 27–29, 290
comb filtering *see* acoustical interference
compression in recordings 85–86

concerts/concert halls: amplified popular music concerts 66–67; critical distance 357; do halls degrade timbral perception? 156–157; draw-away curves 357; electronic acoustical enhancements 242; part of live performances 65
Consumers Union (Consumer Reports magazine) loudspeaker tests 329–334
control rooms *see* recording control and mastering rooms
critical bands *see* hearing
crosstalk: interaural crosstalk cancellation 77–81; in stereo phantom images *see* stereo
cumulative-energy time 346–347
C-weighting curve 320
cylindrical spreading 353

damping factor *see* impedance
dialog intelligibility *see* speech intelligibility
diffusers: depth requirements 283–284; types 283–284, 378–382
diffusion limited by absorption 347–348
dipole loudspeakers 113, 114–116
directivity: directivity index of loudspeakers 53–54, 358; of voices and musical instruments 358
distortion *see* loudspeaker measurements
Dolby Atmos in headphones 472
Dolby Atmos in loudspeakers *see* multichannel audio
Dolby Surround, Stereo and ProLogic *see* multichannel audio
downmixing *see* multichannel audio
draw-away curves: in concert halls 357; in small listening rooms 359–360; for true line source loudspeakers 359–360
dynamic range of recordings 260–261

early reflections *see* loudspeaker measurements
EBU Tech 3276-2nd edition (1998) and 3276-E (2004) 84
Edison “tone tests” 100–101

elevation: the spectral cues *see* localization
envelopment *see* spatial effects
equalization: of anechoic measurements vs. room curves 138–142; can damp resonances in loudspeaker transducers 138; can damp resonances in small rooms 455–461; room curves—results not targets 367–369
equal-loudness contours (Fletcher & Munson curves) *see* loudness
ERBN (equivalent rectangular bandwidth) *see* hearing
ETC (energy-time curve) 185–186

far field of loudspeakers 349–352
Fletcher & Munson curves *see* equal-loudness curves
Floyd Toole biography xix, 9–11, 47–48, 63
flutter echo 283
frequency response *see* loudspeaker measurements, headphones
frequency-weighting curves (A, B, C, etc.) 320
front-of-house (FOH) engineer 67

group delay 22

Haas effect *see* precedence (Haas) effect
Harman International 47–48
harmonic distortion *see* non-linear distortions
headphones: benefits of headphones 473; Berns, headphone survey 508; bone conduction headphones 477; Breebart headphone survey 507; Brüel & Kjær Type 5128 /4620 478–479; case of missing 6dB 471; closed back, open back 476; diffuse-field calibration 490; a distinctive listening experience 158–159; earbuds (no tips) 477; earphones 473; effect of headphone test fixture on measurements 481–484; free-field calibration 490–491; GRAS 43AG/45CA 478–479; Harman headphone research 492–507; Harman headphone target curve 494–496; Harman IE headphone target for B&K 5128 503–506; Harman target tested against other proposed targets 509; head and torso simulators 480; headphone listening test methods 492–494; headphone measurement challenges

480–481; headphone placement and insertion effects 481; headphone test fixtures 478–480; headphone topics for future research 519; headphone type and hearing loss 477–478; headphone type popularity based on sales 475; headphone types 474–477; headphones for spatial audio 518; headphones measured on humans vs. test fixtures 484–486; headphones with headbands 476; headphones without headbands 477; Headphones.com, preference bounds and calibration 515–516; history of preferred headphone calibrations 489–492; HRTF (personalized vs. generalized) 472–474, 487; human anatomical variability 481; Hyperion Project 487–488; Hyperion Sphere 487; IEC 60318–7(2022) 480; IEC 60628–24 (2023) 488; IEC: 60318–4 (2010) 480; immersive audio experience: loudspeakers vs binaural headphones 472; in-ear (IEM) headphones 474; ITU-T P.57 (2021) 480; ITU-T Rec. P. 58(2023) 480; Knowles IE headphone target 510; lack of tactile experience 471; leakage effects 481; Lorho headphone target curve 496; loudspeaker vs. headphone spatial impressions 469; loudspeaker vs. headphone timbral impressions 470; measuring headphone ANC performance and transparency 488–489; Møller et al. diffuse-field targets 496; Multi-dimensional Audio Quality Score (MDAQS) 512; no correlation between headphone price and sound quality 507–509; on-ear (OE) headphones 476; open-ear headphones 477; phantom center experience in headphones 471; physical substitution test method 492; predicting headphone preference based on objective measurements 501–503; preferred levels of bass and treble in headphones 497–499; proper application of Harman headphone research 506–507; reproducibility of headphone measurements 481; reproducibility of measurements on humans 486–487; Rtings headphone target 512–513; segmentation of listeners based on preferred headphone sound profile 499–501; semi-reflective field calibration 491; SenseLab headphone study 513–515; SenseLab headphone testing 489; single stimulus comparisons of headphones 472; SonarWorks headphone target 509–510; SoundGuys headphone target 511; spatial audio target curve 518; summary of Harman target curve versions 516–518; test fixture accuracy in matching

humans 480–481; type 3.3 ear simulator 480; type 4.3 ear simulator 480; virtual headphone method 493–494
hearing: the amplitude and frequency boundaries of what is audible 260–261; audibility of absolute polarity 272; audibility of phase and group delay 269–271; audibility of spectral tilts 260; do we hear waveforms? 270; effects of sound level 248–250 (*see also* loudness); resolution, critical bands, ERBN 267–269; threshold of 260–261; what is heard vs. what is measured 24–26
hearing loss: changes width of critical bands 269; a criterion for subjective evaluations? 122; degrades ability to hear equalization 527; hearing conservation guidelines 248–250, 260–261, 528–530; hidden hearing loss 531–532; modifies judgments of sound quality 7, 35, 122; musicians' earplugs 529; what can audio professionals do? 530–531
home theater design 238–239
HRTF (head-related transfer function) 72–77, 153

imaging: center loudspeaker vs. phantom center 153, 174–175; channel based vs. object based 214–215, 219–221; loudspeaker designs to expand the sweet spot 111–113; stereo soundstage and phantom images *see* stereo
immersive audio *see* multichannel audio
impedance: amplifier output impedance /damping factor 312–317; of loudspeakers 312–313
in-head localization 121
interaural cross correlation coefficients (ICCC or IACC) 199–205, 215
intermodulation distortion *see* non-linear distortion
inverse-square law *see* propagation loss
ITU-R BS.1116–3 (2015) 84

K-weighting curve 320

LEV (listener envelopment) *see* spatial effects
LFE (low-frequency effects) channel 217–218

linear distortions *see* loudspeaker measurements
line-source and truncated line loudspeakers 352–353; CBT (constant beamwidth transducer) 353–356; interaction with floor 353–356; as surround loudspeakers 234–237
listening rooms: acceptable background noise (*see* noise criteria); acoustical treatment suggestions 390–393; diffusion limited by absorption 347–348; effects on loudspeaker sound quality ratings 142–150, 302–303; the Harman positional substitution room 303–306; listener adaptation to rooms 132, 150–151; the original NRCC room 31–32, 303; the range of listening spaces 70; the range of spaces 70; resonances (*see* resonances in rooms); reverberation time (*see* reverberation time)
listening tests/subjective evaluations—loudspeakers: ABX test method 17; auditory memory 19; binaural record/replay methods 142–144, 307–308; blind vs. sighted 288–293, 306–307; controlling the experimental variables 300–324; effects of different rooms and location 142–150, 292, 302–303; effects of hearing loss 35, 122, 292–298; electrical and electronic influences 312–317; equal-loudness testing 308–310, 319–323; evaluations with multiple listeners 306; the final arbiter of quality 4; hearing threshold requirements for listeners 122; listener aptitude, experience and training 35, 288, 292–293, 298, 317–319; male vs. female 293; mono vs. stereo 113–123, 301–302; mono, stereo, multichannel comparison 20, 113–123; multiple loudspeaker evaluations 17, 19, 149–150; MUSHRA test method 17; non-linear subjective rating scales 17–18; paired comparison (A vs. B) evaluations 16, 19; position randomization and positional substitution 303–306; preference/pleasantness vs. accuracy/fidelity 4, 39, 323–324; program choice and presentation 19, 310–312; program influences results 121–122, 149, 292; sighted test problems 15, 288–293; single stimulus listening evaluations 16; sound quality vs. spatial quality 113–123, 196–199; stress 300–301; subjective measurements 8, 15, 287; subjective ratings domestic vs. pro monitor 46; subjective ratings vs. price 46; subjective/objective correlations 59–62, 329–339; test methods compared

16–17; tube power amplifier problem 312–317; wire requirements 312–317
listening through rooms 25, 142–150
listening window *see* loudspeaker measurements
live vs. reproduced demonstrations 102, 104–105
localization of sounds: in-head localization 121; perception of elevation 205–206; reflections affect image position and size *see* reflections
loudness: equal-loudness contours (including Fletcher & Munson curves) 253–255; equal loudness in loudspeaker comparisons 11, 15, 253–255; loudness controls 255–258; measurement of 319–323; phons 253–255
loudspeaker measurements: anechoic methods 23, 27, 36–45, 49–53; ANSI/CTA-2034-A measurement standard 23, 30; changed the industry 56–58; directivity index 53–54; directivity requirements for multichannel systems 225–235; early reflections 51–52; frequency response 22, 246, 251–253; impedance—interaction with tube amplifiers 312–317; inadequate manufacturers' specifications 23, 59–60; linear distortions 263–264; listening window 51; minimum-phase systems 20; non-linear distortion 273–275; NRCC 1986 test results 36–41; Power compression 275–277; resonance detection, evaluation and repair 20, 132–142, 251–253; room curves (*see* room curves); sound power 52–53; sound quality correlates with spinorama data 329–339; spinorama defined and interpreted 49–53, 59–60, 251–253
loudspeakers: 30years of progress 539; 60years of history 541–545; directivity of different designs 113; examples of some subjectively highly-rated loudspeakers 50, 61, 277, 540; monitor loudspeakers vs. consumer loudspeakers 46; stereo image stabilization designs 111–113; voicing 290
LPs 33–34

masking 21–22, 272–275
measurements *see* loudspeaker measurements
microphones 68–69; placement affects sound quality 100–101

minimum phase: in loudspeaker transducers 20; in low-frequency room resonances 455–458
monitor loudspeakers: custom “design” by equalization 89–90, 387; a digression in pro audio monitoring 88–89; a factor in the circle of confusion 28; a factor in the translation of recordings 86; historical and modern examples 87–89
mono vs. stereo listening 16, 113–123
movie sound *see* cinema sound
multichannel audio: arrangements in rooms 152, 195–197, 199–205, 222; bass management 217–219; center loudspeaker designs 229–231; center loudspeaker vs. phantom center image 153, 174–175; channel based vs. object based 214–215, 219–221; channel numbering schemes 221; Dolby Atmos calibration recommendations 223–224; Dolby Surround, Stereo and ProLogic 213; elevation/height 205–206; envelopment—how many channels are necessary 199–205; a factor in dialog clarity and enjoyment 532–534; immersion/Atmos 216, 219–221, 223–224, 242–243; LFE (low frequency effects) channel 217; loudspeaker directivity requirements 224–237; quadraphonics 199–200, 211–212; ray diagrams in rooms 152–153, 195–196, 214; recorded “spaces” reduce audibility of resonances in loudspeakers 154–156; surround loudspeaker designs 231–237; upmixing and downmixing 216, 241–242
MUSHRA test method 17
musicians’ earplugs 529

near field monitoring (close listening) 352
near field of sound sources 349
next-generation audio 243–244
noise criteria (NC): acceptable levels 262–263; curves 260–263
non-linear distortions 21, 272–275
non-linear subjective rating scales 17–18
NRCC (National Research Council of Canada) 11

occupational hearing conservation guidelines 248–250, 529–530

Olive's subjective/objective correlations 329–339
operational room response (ORR) *see* room curves
opinions 3–4, 15

paired comparison (A vs. B) subjective evaluations 16
particle velocity: in relation to resistive absorbers 31; in room resonances/modes 546
phase *see* hearing
phons *see* loudness
placebo effect 4
point source 348; effect of solid angle 397–400
polarity *see* phase
power amplifier: audibility of clipping 249, 275, 309; the damping factor deception 418; most tube amplifiers change the sound of loudspeakers 5; output impedance/damping factor 314
precedence (Haas) effect: the fusion interval differs for different sounds 175–177; the fusion interval in the real world 169–174; the fusion interval for speech 164–165, 168; the original 1949 Haas experiments 163–167
preference vs. accuracy *see* listening tests
presbycusis 528
propagation loss: air absorption 370–372; inverse-square law 349; screen loss 372; in surround loudspeakers 234–237
psychoacoustics 8, 63

Q and the audibility of resonances *see* resonances
quadraphonics 211–212

recording control and mastering rooms: art should be decided by opinion, technology by measurements 90; blackbird studio C 388–390; different "fashions" at different times 83–86; hearing may be a factor in what is popular 530; home studios 83, 85; where the art is created and judged 83; *see also* monitor loudspeakers

reflections: acoustical treatment options 177–180; alleviate the audibility of stereo crosstalk comb filtering 109–111; audible effects at different directions, delays and levels 167–169, 171–172; audible effects for different sounds 175–177; can improve speech intelligibility 111; depend on loudspeaker directivity 113, 116; depend on surface treatment 372–373; early reflections curve in the spinorama 51–53; effects on real and phantom sound images 174–175; floor reflections 353–357; measuring reflections in rooms 185–186; modify the precedence effect 169–174; ray diagrams for different playback formats 51, 152; subjectively preferred levels for different programs 198; subjectively preferred times, levels and directions 189–196
reproducing "reality": binaural techniques 71–74; reconstructing sound fields 71; reproducing essential cues 81
resonances in headphones 159
resonances in loudspeakers: attenuating resonances in loudspeaker transducers 138–142; audibility as a function of Q, frequency and program 132–142; audibility influenced by playback format: mono, stereo, multichannel 113–123, 141–142, 155–156; do we hear the peaks or the ringing? 134–136; identifying resonances in measurements 39, 132–142, 265–267; the "knuckle test" 21; mechanical and acoustical resonances 21; peaks vs. dips in frequency response 138; resonances are poorly revealed in room curves 139–140; why are resonances important? 20, 264
resonances in rooms (low frequency) *see* room resonances
resonances in voices and musical instruments 20, 131
reverberation time (RT): calculations using Sabine equation 345, 393; of cinemas and screening rooms 343; the contribution of various room furnishings and finishings 343; IEC, ITU and CEDIA recommendations 343–344; optimum range for speech intelligibility 344; is there a more useful metric? 346–347; of typical consumer and custom designed listening rooms 341–345
reviewing loudspeakers 4, 60, 299–300
RLB weighting curve 320

room as an experimental variable 142–150
room curves: for bipole (almost omni) loudspeaker 364; Harman curve: a result not a target 367–369; JBL Pro M2 measured in numerous rooms 364–367; meaning of 360–361; predicted from anechoic data 41–45; predicted from spinorama data 360–364
room equalization/calibration: results compared to anechoic loudspeaker equalization 400; room curves are results, not targets, but useful at low frequencies 23–24, 44–45
room mode/resonance attenuation and manipulation 418–420, 428–429; absorption/bass traps 429–434; do we hear the peaks or the ringing? 461–464; employ two subwoofers 438–440; global equalization affects both amplitude and time responses 455–461; multiple subwoofers (active methods rectangular rooms 448–450; advanced passive methods 444–448; basic passive methods 441–444; sound-field management (SFM) for any room 450–455); multiple subwoofer performance metrics (MSV, MOL, VSA) 441–442, 448; optimize listener location 419, 434–437; optimize subwoofer location 419, 436–437
room modes/resonances/standing waves: dominant below the transition frequency 128–130; effects of loudspeaker and listener location 130; equation for calculation 416; non-rectangular rooms 426–427; resonant modes displayed 130, 416, 417; room pressurization mode/room gain 420; standing waves displayed 130, 417, 432; standing waves explained 280–282; is there an ideal room? 424–426; three basic problems 427; types 414–415

Sabine, Wallace Clement 65
Schroeder frequency *see* transition frequency
scientific method applied to audio 7, 71
screen loss *see* propagation loss
Sean Olive, biography xix, 54
single stimulus listening evaluations 16
Sound Field Management (SFM) 450–455
sound power *see* loudspeaker measurements

sound pressure level (SPL) 247; of common sounds 248–250; vs. distance in rooms—draw-away curves (related to amplifier power 248–250; related to subjective effects and hearing conservation 248–250); vs. sound level 247–248
sound quality: audibility of spectral tilts 260; dynamic range of recordings 260–261; effects of loudspeaker directivity 278; importance of bass quality and bass extension 36–41, 333–335; importance of direct sound 335–338, 369–371; judged in mono, stereo, multichannel 113–123, 155–156, 532–534; methods for subjective evaluation (*see* listening tests); Olive's subjective/objective correlations 329–339; playback sound level is a factor (equal loudness contours) 253–255; Price vs. sound quality in loudspeakers 46; ranks with spatial quality in importance 113–123; resonances are a dominant factor (*see* resonances); spectra of music 260–261; the spinorama contains much of the information 138–141; tone controls are useful 258–260; voicing loudspeakers 290
sound reproduction objectives 97–99
soundstage and imaging *see* stereo
spatial effects: ASW apparent source width 98; distance 98; influenced by playback level 99; influenced by reflection direction, timing and level 189–196; in-head localization 121; LEV listener envelopment 99, 199–205; in monophonic listening 113–123; rank with sound quality in importance 113–123, 196–199; stereo bass/low-frequency envelopment 464–465
specifications: what we get vs. what we need 59–60
speech intelligibility: beneficial effects of early reflections 151; degraded by multiple active channels 532–554; degraded in phantom center image by acoustical crosstalk 111; improved by adaptation to rooms 150–151
speed of sound 277
spherical spreading 349
spinorama *see* loudspeaker measurements
standing waves *see* room modes/resonances
steady-state room curves *see* room curves

stereo: effect on the audibility of loudspeaker resonances 154–156; image stabilizing loudspeakers 111–113; origins of 103–104; phantom image problems 108–111, 153; phantom image vs. real loudspeaker 152, 174–175; soundstage and imaging 23, 105–108, 113–123, 152–153; stereo bass, low-frequency envelopment 464–465
subjective evaluations *see* listening tests
subwoofers *see* bass management, room modes/resonances
summing localization 164–165
surround channels: acoustical treatments can help 237–238; delivering localized and non-localized sounds 214; directivity requirements 227–228; improving surround effects using special loudspeakers 234–237; problematic "dipoles" 231–237

Todd Welti biography xix, 54–55
tone, tilt and loudness controls 258–260
transition frequency 128–130
translation of recordings into the real world: circle of confusion 28; custom "design" loudspeakers by equalization 89–90, 387; historical perspective 383–385; lack of standards is a problem 68, 86–90; recommendations based on science 386–388; the reproduction/manufacturer's perspective 385–386; voicing loudspeakers 290

upmixing *see* multichannel audio

virtual reality headsets 77
voicing loudspeakers 290

waterfall diagrams 265–267
waveforms: do we hear them? 270
wavelength: the basis of room resonances 280–282; as a function of frequency 277–278; related to acoustical absorbers and diffusers 282–284; related to loudspeaker directivity 278–279
weakest link in audio 56
website address 6

wire *see* listening tests

X-curve in cinema sound 91